

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

RÔLE DE L'ADN MITOCHONDRIAL DANS L'ADAPTATION AU CLIMAT

MÉMOIRE PRÉSENTÉ
COMME EXIGENCE PARTIELLE DE LA
MAÎTRISE EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

PAR
YANNICK NOËL

JUIN 2020

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

*“Knowledge is only part of the fight
for survival.”*

Max Brooks

REMERCIEMENTS

Je voudrais remercier mon directeur de recherche, le Dr Emmanuel Milot, qui m'a permis de découvrir le vaste domaine de la biologie évolutive humaine en me donnant l'opportunité de réaliser mon projet de maîtrise. Ces conseils, son support et son expertise ont été très appréciés. Je tiens aussi à remercier toute l'équipe ainsi que les personnes qui ont séjourné au laboratoire de génétique des populations au cours de ma maîtrise. En particulier, Alexandra Doyon qui m'a guidé lors de mes premiers pas avec le logiciel R. Walid Mawass qui a su répondre à de nombreux questionnements concernant mon projet ainsi qu'à Carla Aimée-Jubin pour son aide dans l'élaboration des modèles statistiques.

De plus, je tiens à remercier nos collaborateurs Damian Labuda et Claudia Moreau du CHU Sainte-Justine, l'équipe de BALSAC à Chicoutimi et Catherine Laprise de l'UQAC pour le partage de données, leur disponibilité et les informations qu'ils m'ont prodigué.

Merci aux Fonds de recherche du Québec – Nature et technologie et le Conseil de recherches en sciences naturelles et en génie du Canada qui ont permis le financement de ce projet.

Je désire également exprimer mon appréciation envers la famille Lygitsakos, mes employeurs de longue date, pour leur compréhension et leur flexibilité m'ayant permis de faire passer mes études en priorité. Finalement, je suis énormément reconnaissant envers ma conjointe Sara, mère de mes adorables petits jumeaux Éliott et Zack. Elle a su alimenter ma motivation, m'épauler et m'encourager tout au long de mon cheminement.

RÉSUMÉ

Les variations dans l'ADN mitochondrial (ADNmt) ont le potentiel de causer toutes sortes de désordres métaboliques et physiologiques. Plus de 600 mutations sont associées de près ou de loin à des pathologies humaines. Il est maintenant bien établi que certains des variants mitochondriaux (haplotypes ou mitotypes) ont la capacité d'affecter le fitness (valeur sélective darwinienne) d'un individu, notamment en altérant la chaîne de transport d'électrons, conduisant ainsi à des perturbations au niveau de la phosphorylation oxydative. Comme les gènes mitochondriaux ont un rôle central dans le métabolisme, il est donc fortement possible qu'ils aient été soumis intensément à la sélection naturelle.

C'est d'ailleurs ce qu'indiquent bon nombre de recherches ayant détecté des traces de sélection dans le génome mitochondrial. De plus, cette sélection a été mise en lien avec le climat. Une des hypothèses stipule que l'ADNmt joue un rôle dans l'adaptation au climat et que ce dernier aurait façonné la répartition des variants mitochondriaux dans le monde. Bien que les preuves s'accumulent, les études réalisées à ce sujet ont utilisé des méthodes statistiques indirectes soumises à certains biais. Ces biais proviennent principalement des processus démographiques qui miment la sélection et qui sont très difficiles à séparer de la mesure de la sélection. C'est pourquoi des chercheurs remettent en cause la valeur des résultats scientifiques pour soutenir cette hypothèse.

De plus, la fonctionnalité d'un gène ne peut être pleinement évaluée que dans l'environnement dans lequel il s'exprime. C'est pourquoi j'utilise ici une approche différente et originale pour tester « en temps réel » l'hypothèse climatique dans la population préindustrielle canadienne-française du Québec. En utilisant l'imputation généalogique des haplotypes, nous pouvons associer la variation dans l'ADNmt à des traits d'histoire de vie (THV) individuels au cours des décennies qui ont suivi la fondation de cette population. Ce qui nous permet de mesurer la sélection sur des mitotypes génération par génération et sa relation avec les variables climatiques, en nous basant sur des séries chronologiques d'indicateurs climatiques de la sévérité de la température, de l'hiver et de la sécheresse. La fondation de la Nouvelle France, suivie de la colonisation de nouveaux territoires donnant naissance à des sous-populations telles que le Saguenay-Lac-Saint-Jean, ont exposé les colons à une véritable expérience naturelle : les génomes européens, présumément adaptés à un climat plus clément, ont été subitement exposés à de nouvelles conditions.

Nous montrons, à partir de la mortalité infantile, qu'une sélection sur l'ADNmt causée par des variations climatiques est bel et bien possible dans une population humaine. Nos résultats montrent une forte association entre la probabilité de mortalité infantile et le mitotype des individus. De plus, la probabilité de mortalité propre à chacun des mitotypes varie selon le climat environnant la naissance.

Mots-clés : ADNmt, sélection naturelle, climat, mortalité infantile.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ.....	iv
LISTE DES FIGURES ET TABLEAUX.....	vii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	viii
CHAPITRE I	
INTRODUCTION.....	1
1.1 La sélection naturelle.....	1
1.2 Détecter la sélection naturelle.....	2
1.3 Limite des tests moléculaires.....	3
1.4 ADN mitochondrial	5
1.5 La chaîne de transport d'électrons.....	7
1.6 ADN mitochondrial et adaptation.....	10
1.7 Problématique.....	13
1.8 Mécanismes possibles d'adaptation.....	14
1.9 La mortalité infantile comme mesure de la sélection mitochondriale.....	17
1.10 La mortalité infantile	20
1.11 Objectif de la recherche	22
1.12 Approche expérimentale.....	22
CHAPITRE II	
MTDNA AND CLIMATE ADAPTATION: A DIRECT TEST OF SELECTION USING MOLECULAR AND GENEALOGICAL DATA	25
2.1 Contribution des auteurs.....	25
2.2 Résumé de l'article	26
2.3 Article complet en anglais: mtDNA and climate adaptation: a direct test of selection using molecular and genealogical data.....	27
Abstract.....	27
Introduction.....	28
Methods	30
Data and study population	30

Molecular data	34
Probabilistic imputation of mtDNA	34
Climatic variables	36
Statistical analysis.....	37
Results	39
Climate effect on infant mortality	39
mtDNA haplogroups analysis.....	45
mtDNA haplotype analysis.....	49
Discussion.....	52
Conclusion	55
Supplementary material	56
References.....	59
CHAPITRE III	
DISCUSSION ET PERSPECTIVE	66
RÉFÉRENCES BIBLIOGRAPHIQUES.....	76

LISTE DES FIGURES ET TABLEAUX

Figure		Page
1.1	Exemple d'un goulot d'étranglement (A) et d'un effet fondateur (B).....	4
1.2	Schéma de l'ADNmt.....	6
1.3	Schéma de la chaîne de transport d'électrons	8
1.4	Représentation de la fuite de protons.....	9
3.1	Distribution de la fécondité dans la population du Saguenay-Lac-Saint-Jean.....	69
3.2	Manhattan plot circulaire de l'association des SNPs mitochondriaux avec la mortalité infantile au Saguenay-Lac-Saint-Jean	73
3.3	Différentes trajectoires de fitness temporelle pour 4 marqueurs d'ADNmt de la région de contrôle 1, de 1775 à 1915	75
Tableau		
1.1	Pathologies mitochondriales transmises maternellement et s'exprimant parfois dans l'enfance	19
1.2	Distribution de la mortalité infantile dans différentes populations préindustrielles.....	21

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ADN	Acide désoxyribonucléique
ADNmt	ADN mitochondrial
ADP	Adénosine diphosphate
ARN	Acide ribonucléique
ARNr	ARN ribosomal
ARNt	ARN de transfert
ATP	Adénosine triphosphate
ATP/O	Ratio ATP-Oxygène
TAB	Tissus adipeux brun (brown adipose tissue)
dN/dS	Ratio de mutations non-synonyme/synonyme
ETC	Chaîne de transport d'électrons (electron transport chain)
IM	Infant mortality
P:C	Ratio protéines carbohydate
RCR	Ratio de contrôle respiratoire (respiratory control ratio)
RFLP	Polymorphisme de longueur des fragments de restriction (<i>restriction fragment length polymorphism</i>)
ROS	Espèces réactives de l'oxygène (reactive oxygen species)
SLSJ	Saguenay-Lac-Saint-Jean
SNP	Polymorphisme nucléotidique (<i>single-nucleotide polymorphism</i>)
THV	Traits d'histoire de vie

CHAPITRE I

INTRODUCTION

1.1 La sélection naturelle

La sélection naturelle est probablement le mécanisme le plus connu de l'évolution, les autres étant la dérive génétique (phénomènes aléatoires), les mutations et la migration (flux de gènes). Elle permet l'adaptation d'une population à son milieu en métamorphose constante. De façon simple, la sélection naturelle opère lorsqu'il y a de la variabilité dans les traits chez les individus d'une population et que cette variabilité conduit à une différence de fitness en eux (c.-à-d. leur succès reproducteur ou valeur sélective) dans un environnement donné. Le fitness est déterminé par la survie, l'appariement sexuel et la reproduction de l'individu. Les individus possédant des traits leur conférant un plus grand fitness vont avoir plus de descendants que les autres et seront sélectionnés de façon positive par rapport aux individus ne possédant pas ces traits avantageux (ces derniers seront sélectionnés négativement). Les traits héréditaires seront donc transmis à leur progéniture et il y aura par le fait même une augmentation de la fréquence des valeurs de traits favorables, accompagnée d'une diminution des traits désavantageux dans cette population. Ce qui conduira à une plus grande proportion de la population ayant les traits adaptés à cet environnement (Charlesworth et al., 2010).

L'expression de ces traits étant en partie d'origine génétique, les fréquences alléliques des gènes varieront nécessairement en fonction de la sélection. Une sélection positive augmentera la fréquence d'un allèle, pouvant même conduire à sa fixation (toute la population est porteuse de cet allèle) alors qu'une sélection négative diminuera sa fréquence et peut même causer à sa disparition. Ce qui dicte la vitesse à laquelle se s'effectue ces changements est la force de la sélection et la balance entre la sélection et la dérive génétique (Hartl et al., 2007; Templeton, 2006).

1.2 Détecter la sélection naturelle

De nombreux types de traits s'expriment chez un individu : comportementaux, morphologiques, physiologiques, reproducteurs, etc. Lorsque les traits sont quantifiables, il est possible de détecter et de mesurer la force de la sélection s'exerçant sur eux en mesurant la covariance entre les traits et le fitness des individus. Le fitness étant le succès reproducteur, il se mesure facilement sur des organismes avec un temps de génération court. Chez des organismes avec un temps de génération plus long, il est commun de diviser le fitness en composantes telles que la survie à la reproduction, l'âge à la première reproduction et la fécondité. Ces composantes du fitness sont communément appelées traits d'histoire de vie (THV). Par exemple, si un trait réduit les chances de survie d'un organisme jusqu'à la reproduction, les porteurs de ce trait auront moins de chance en moyenne de trouver un partenaire et se reproduire. Par conséquent, le fitness des porteurs en sera réduit.

Il est aussi possible de détecter la sélection survenue par le passé à partir de la signature laissée par les mutations dans la séquence d'ADN. Il est toutefois plus difficile de quantifier la force qu'elle exerçait. De nombreux tests statistiques ont été développés à cette fin. Ils se distinguent principalement par l'échelle à laquelle ils sont appliqués, c'est-à-dire, par des comparaisons entre espèces, entre populations d'une même espèce ou entre les individus d'une même population. Parmi ceux-ci, le test de ratio dN/dS (Nei et al., 1986), le test de McDonald-Kreitman (McDonald et al., 1991) et le test de Tajima (Tajima, 1989) compte parmi les plus utilisés (Vitti et al., 2013).

Le test de ratio dN/dS utilise des séquences d'espèces divergentes et compare le taux de substitutions non-synonymes sur la séquence d'ADN (N , conduisant à un changement d'acide aminé dans la protéine) à celui de substitutions synonymes (S , pas de changement d'acide aminé) en utilisant le codon comme unité d'analyse. Différents modèles existent pour ce faire, avec différents niveaux de complexité mathématique pour tenir compte, par exemple, des substitutions successives sur un même site. Le modèle de Nei-Gojobori (1986), par exemple, ne prend pas en compte le ratio de transition/transversion et assume que les deux types de mutations surviennent à la même fréquence. D'autres modèles

estiment le taux de substitution selon la position sur le codon (Goldman et al., 1994; Yang et al., 2000). Les régions d'un gène sous sélection purifiante (régions protéiques conservées) démontreront un excès de mutations synonymes, ce qui donnera un ratio dN/dS plus petit que 1 alors qu'un ratio plus grand que 1 indiquera une tendance à long terme de sélection positive.

Le test de McDonald-Kreitman compare la variation génétique au sein d'une espèce à la différence inter-espèces en utilisant un ratio de mutations non-synonymes et synonymes mesurés sur un gène complet. Sous l'hypothèse nulle de l'absence de sélection lors de la divergence des deux espèces, on s'attend à ce que le ratio dN/dS au sein d'une espèce soit équivalent à celui entre les espèces (divergence due essentiellement à la dérive génétique).

Finalement, le test de Tajima mesure avec la statistique D la différence entre deux estimations de diversité génétique dans un échantillon de séquences de nucléotides. La première estimation est obtenue en calculant le nombre de sites polymorphiques total de l'échantillon. La deuxième estimation se fait en calculant la moyenne des différences polymorphiques par paires de gènes. L'hypothèse nulle stipule que ces deux estimations sont équivalentes lorsque l'évolution est neutre.

1.3 Limite des tests moléculaires

Les méthodes utilisées pour détecter la sélection à partir de la séquence d'ADN demeurent sujettes à une interprétation prudente. Le test de Tajima, par exemple, repose sur des prémisses qui sont rarement respectées par les populations naturelles. Il considère une taille de population constante, aucune migration ainsi qu'un appariement aléatoire des partenaires de reproduction (*panmixie ou random mating*). Plusieurs phénomènes peuvent aussi faire varier la fréquence des allèles dans une population et être confondus avec la sélection, ce qui peut conduire à des inférences erronées. C'est entre autres le cas de changements démographiques rapides qui peuvent changer le pool génétique et laisser dans les génomes une signature ressemblant à celle de la sélection (Stajich et al., 2005).

Par exemple, un goulot d'étranglement et un effet fondateur (Figure 1.1), deux phénomènes conduisant à une réduction de la diversité génétique qui peuvent être confondus avec la fixation d'une mutation bénéfique par sélection positive. Lors de la fixation, il peut également y avoir une réduction de la diversité allélique par « *hitchhiking* » en entraînant des gènes qui ne sont pas sous sélection, mais qui sont physiquement liés à celui sélectionné (Hung et al., 2014). Théoriquement, au niveau moléculaire, on s'attend à ce que les effets démographiques se reflètent sur l'ensemble du génome alors que la sélection, elle, laisse une trace sur un ou des segments précis de l'ADN. Cependant, il demeure compliqué de séparer démographie et sélection par ces approches (Ramírez-Soriano et al., 2008).

De plus, ces tests de neutralité sont aussi sensibles à la taille de l'échantillon. Si le test est effectué à partir d'un échantillon trop petit, le nombre de mutations dans la population sera sous-estimé, car il ne prendra pas bien en compte les allèles rares. D'ailleurs Subramanian (2016) montre qu'en utilisant les exomes de 512 humains au lieu de 16, la détection de mutations jugées adaptives passe de 24 % à 10 %.

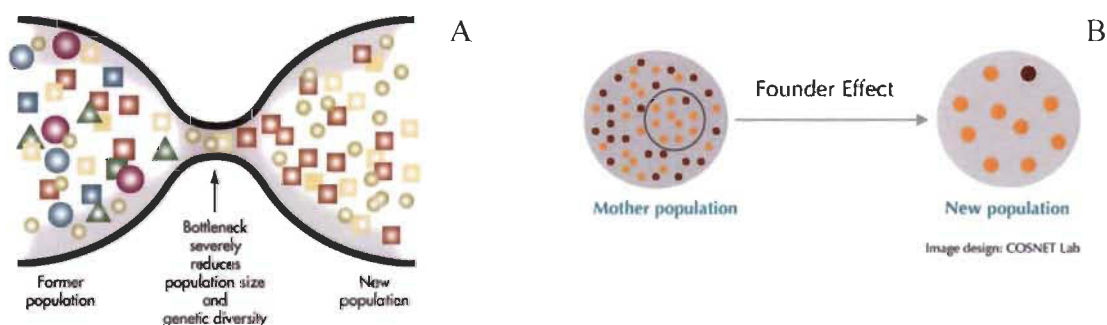


Figure 1.1 Exemple d'un goulot d'étranglement (A) et d'un effet fondateur (B).
 Un événement fondateur se produit lorsqu'une partie d'une population colonise un nouvel environnement, tandis que le goulot d'étranglement peut survenir lorsqu'une population fait face à un événement dramatique, réduisant subitement sa taille. Dans les deux cas, la diversité génétique sera réduite et les polymorphismes les plus communs auront plus de chances de persister.
 Image provenant de : <https://thenaturalhistorian.com>

1.4 ADN mitochondrial

La mitochondrie provient très probablement de l'intégration par endocytose d'une bactérie proche du genre *Rickettsia* par une cellule eucaryote il y a de ça environ deux milliards d'années (Kurland et al., 2000). Au fil du temps, son génome se serait intégré dans l'ADN nucléaire, de sorte que l'ADNmt aurait lui-même diminué progressivement de taille. À ce jour, il ne reste qu'un plasmide double brin de 16 à 20 kb avec un code génétique qui est en partie différent du code universel de l'ADN nucléaire. Ce dernier diffère aussi selon l'espèce (exemple : AGA et AGG sont des codons stop chez l'humain, sérine chez *Drosophila* et non des codons arginine comme sur l'ADN nucléaire). L'activité de la mitochondrie est hautement régulée par le noyau. La majorité des protéines s'y trouvant sont issues des gènes nucléaires. Cependant, l'ADN mitochondrial encode toujours 22 ARN de transfert, deux ARN ribosomiaux ainsi que 13 protéines (Figure 1.2). Ces 13 protéines sont transcrites uniquement à l'intérieur de la mitochondrie par les ribosomes mitochondriaux en utilisant son propre ARN de transfert. L'ensemble de ces 13 protéines sont impliquées dans la formation de la chaîne de transport d'électrons (ETC); (Figure 1.3), qui est une composante essentielle de la respiration cellulaire, un processus convertissant l'énergie contenue dans les nutriments en ATP, principale source d'énergie de la cellule (Lodish, 1997).

Une cellule peut contenir des milliers de copies d'ADNmt selon la demande énergétique du tissu dans laquelle elle se trouve. On retrouve généralement de 2 à 10 copies d'ADN par mitochondrie et un nombre de mitochondries allant de 100 à 1000 par cellule (Fazzini et al., 2018). Lors de sa réplication, le taux de mutation de l'ADNmt est environ 100 fois plus élevé que celui de l'ADN nucléaire. Ceci est dû en partie à un mécanisme de réparation moins performant que dans le noyau. Il a aussi été suggéré que l'enzyme ADN polymérase gamma, codé par le noyau, a tendance à répliquer moins fidèlement l'ADNmt en raison d'une plus grande concentration de désoxyguanosine triphosphate (un précurseur de l'ADN) présent dans la mitochondrie (Song et al., 2005). Elle intégrerait donc davantage de mauvais nucléotides. Les taux de mutations varient selon la partie du génome mitochondrial. La région de contrôle, lieu de l'origine de réplication, est connue pour avoir des régions hypervariables, dont le taux de mutation est

estimé à 4.3×10^{-3} par nucléotide par génération, tandis que le taux pour la région de contrôle au complet est estimé à 3.5×10^{-5} . Ce taux diminue à 8.8×10^{-6} par nucléotide par génération pour les régions codantes (Santos et al., 2008).

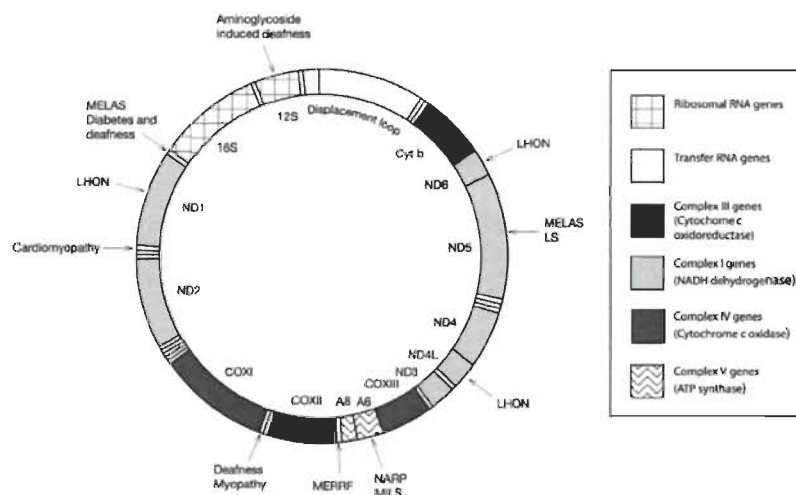


Figure 1.2 Schéma de l'ADNmt.

Les gènes codants pour les protéines sont indiqués à l'intérieur du cercle et certaines pathologies associées aux polymorphismes à l'extérieur. ND : NADH déshydrogénase, Cytb : cytochrome b, COX : cytochrome c oxydase, ATP : ATP synthase. Les complexes de l'ETC et les gènes mitochondriaux codant pour des sous-unités de ces complexes sont :

- Complexe I : ND1, ND2, ND3, ND4L, ND4, ND5 et ND6
- Complexe III : Cytb
- Complexe IV : COXI, COXII et COXIII
- Complexe V : ATP6 et ATP8

Image provenant de : Chapman et al. (2014).

Chez la plupart des espèces dont l'humain, la transmission de l'ADNmt se fait uniquement par la mère, mis à part de rares cas de « fuites paternelles » (Lutz-Bonengel et al., 2019) et ne subit pas de recombinaison. Les mutations survenant dans les cellules germinales, susceptibles d'être transmises à la prochaine génération, font l'objet d'une forte sélection purifiante durant l'ovogenèse (De Fanti et al., 2017). Ce sont donc les mutations neutres ou légèrement délétères qui seraient majoritairement transmises aux descendants. Malgré tout, certaines mutations associées à des phénotypes pathologiques importants peuvent être transmises. Leur effet délétère peut par exemple s'exprimer après l'ovogenèse ou affecter peu le fitness de l'individu (une maladie n'implique pas

automatiquement qu'un individu aura une plus petite descendance que la moyenne). À ce jour, on connaît plus de 600 mutations mitochondriales associées de près ou de loin avec une maladie humaine (MITOMAP: A Human Mitochondrial Genome Database. <https://www.mitomap.org>, 2008).

Dans la littérature, les différentes séquences d'ADNmt sont classifiées par haplogroupes selon les mutations qu'elles arborent. Un haplogroupe se définit comme un ensemble de séquences d'ADNmt avec des polymorphismes communs conservés phylogénétiquement (Weissensteiner et al., 2016). Pour qu'une séquence soit classifiée dans un groupe particulier, elle doit présenter une liste de mutations bien précises. Pour chacune des mutations successives qui seront conservées, on retrouvera une déclinaison de la classification. Par exemple, pour l'haplogroupe H, on trouvera les sous-haplogroupes H1, H1a et H1b, etc. Chacun de ces sous-haplogroupes consiste en une séquence d'ADNmt unique. Nous ferons références dans ce texte au terme haplotypes pour désigner chacune des séquences uniques de l'ADNmt et au terme haplogroupe pour désigner les groupes de séquences appartenant à la première lettre.

Il y a de fortes raisons de croire que les polymorphismes définissant ces haplogroupes soient physiologiquement importants, car les haplogroupes mitochondriaux ont été associés à un bon nombre de maladies comme l'Alzheimer (Ridge et al., 2018), le Parkinson (Swerdlow, 2012), le diabète de type 2 (Charoute et al., 2018) et plusieurs cancers (Singh et al., 2009).

1.5 La chaîne de transport d'électrons

La respiration cellulaire comprend tous les processus conduisant à la formation d'ATP et englobe la glycolyse, le cycle de Krebs, la bêta oxydation des lipides et la dégradation des protéines. Elle peut se faire en présence d'oxygène (respiration aérobie) ou son absence (glycolyse anaérobie). Lors du catabolisme des nutriments, les glucides, acides gras et protéines sont métabolisés et passent par le cycle de Krebs, où il y a formation d'une variété de molécules dont le NADH, le FADH₂ et le succinate.

Ces métabolites sont amenés à la chaîne de transport d'électrons pour exécuter la phosphorylation oxydative, la dernière étape de la respiration cellulaire aérobie.

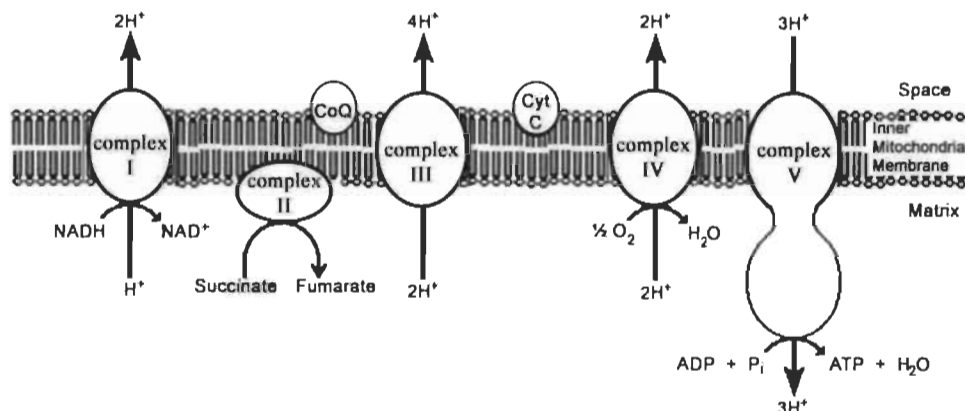


Figure 1.3 Schéma de la chaîne de transport d'électrons.

Image provenant de : <https://cellcode.us>

Le complexe I de l'ETC catalyse l'oxydation du NADH. Le complexe II catalyse l'oxydation du FADH₂ libre, mais peut aussi faire le transfert des électrons du succinate au FAD à l'aide de l'enzyme succinate déshydrogénase pour générer du FADH₂ et ensuite catalyser son oxydation. La coenzyme Q elle sert de pont en acceptant les électrons des complexes I et II pour les transférer au complexe III qui ira ensuite vers le complexe IV. De là, ils seront finalement acceptés par des molécules d'oxygène. Tout au long de ce processus, les complexes I, III et IV agissent comme pompes à protons, envoyant les H⁺ à l'extérieur de la membrane interne mitochondriale, créant un gradient de charge positive. Ces protons sont retournés à l'intérieur via le complexe V, permettant la génération d'ATP par l'ATP synthase (Cooper, 2000).

Le gradient de protons généré pendant la phosphorylation oxydative n'est pas parfaitement couplé à la production d'ATP. C'est-à-dire qu'une bonne partie de ces protons est retournée à l'intérieur de la membrane, soit par diffusion, soit par des protéines découplantes (UPCs; Figure 1.4) et ne servent pas à la production d'ATP. L'énergie libérée lors de la dégradation des nutriments ne servant pas à la production d'ATP sera libérée sous forme de chaleur (thermogénèse). La simple diffusion des protons

contribue d'ailleurs de façon importante au taux métabolique de base (quantité d'énergie dépensée au repos), car le processus de la phosphorylation oxydative s'effectue en continu afin d'avoir une alimentation constante en ATP dans les cellules. L'organisme dépense donc constamment de l'énergie pour produire de l'ATP, mais une bonne partie est relarguée sous forme de chaleur (Jastroch et al., 2010). Les UPCs quant à eux comprennent plusieurs isoformes. L'UPC1 se retrouve principalement dans le tissu adipeux brun et contribue à la thermogénèse. Elle est régulée par la stimulation des adipocytes qui active les voies de signalisation liées à l'AMPc, ce qui engendre l'expression de cette protéine et la libération de ses activateurs comme les acides gras libres. On ne sait pas encore exactement ce qui active les autres UPCs, mais certaines recherches suggèrent un rôle pour les acides gras libres et certains radicaux libres (Rousset et al., 2004).

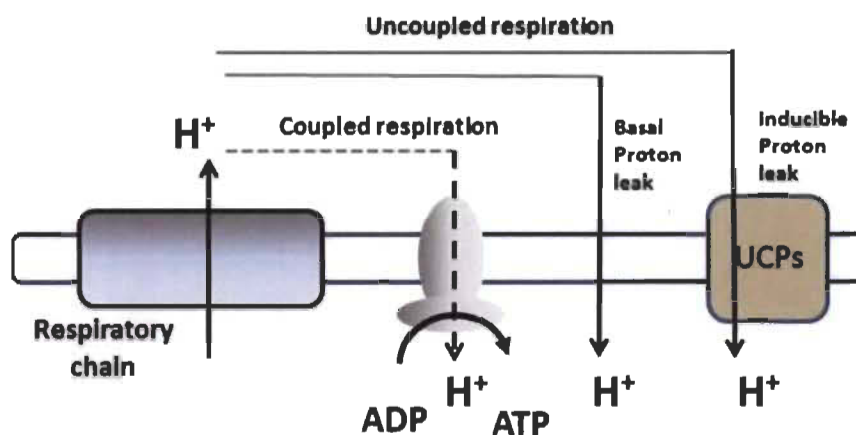


Figure 1.4 Représentation de la fuite de protons.
Image provenant de : Busiello et al. (2015).

Toutefois, dans la plupart des types cellulaires la majeure partie de l'oxygène consommé sert à produire de l'ATP (Zhang et al., 2012). L'efficacité et la stabilité l'ETC sont variables et répondent aux changements dans l'environnement cellulaire par l'entremise des dynamiques mitochondriales comme la fission et la fusion des mitochondries et des modifications structurelles de la membrane interne, notamment le remodelage des crêtes mitochondriales. Les complexes de l'ETC sont situés le long des crêtes et des dimères de l'ATP synthase maintiennent les crêtes courbées. Des modifications dans la structure des crêtes peuvent survenir avec des changements d'état énergétique de la cellule ou lors de stress cellulaires. Par exemple, pour répondre à

un besoin énergétique, une contraction des crêtes permet de libérer l'ATP synthase et augmenter la production d'ATP. Dans le cas de famine cellulaire par exemple, l'élongation de la mitochondrie et l'assemblage de l'ETC en super complexes respiratoires permet la promotion de la respiration cellulaire (Baker et al., 2019).

L'ETC permet non seulement la production d'ATP, mais est aussi la source principale de la production des réactifs de l'oxygène (ROS), principalement l'anion superoxyde. À plusieurs étapes de l'ETC, plus particulièrement dans les complexes I et III, il y a des possibilités de fuite des électrons (Murphy, 2009). Ces électrons vont réduire l'O₂ en O₂•-. Le taux de production de cet anion est d'ailleurs directement lié à la force motrice du gradient de proton (Granger et al., 2015). La production de ROS, tout comme la production d'ATP, est dépendante de l'environnement cellulaire et répond aux mêmes changements de conformation de l'ETC qui sont induit par les modifications des structures mitochondriales, l'assemblage en super complexes et plusieurs autres facteurs. Les ROS sont d'importants signaux cellulaires, mais en trop grande quantité ils sont aussi dommageables pour la cellule, car ils sont très réactifs et peuvent endommager les structures cellulaires comme les membranes, les lipides, les protéines et les acides nucléiques pouvant ainsi entraîner le vieillissement cellulaire ou la mort de la cellule (Bergamini et al., 2004). On verra plus loin que l'importance de ce stress oxydatif peut être associée à des traits d'histoire de vie comme la longévité.

1.6 ADN mitochondrial et adaptation

Depuis quelques années, avec la venue du séquençage de nouvelle génération, il est beaucoup plus abordable et rapide de générer des données à l'échelle génomique. L'un des champs d'application importants utilisant le séquençage de masse mis à disposition par ces technologies consiste à savoir si la sélection naturelle opère encore chez l'humain, ce qui aide à comprendre l'évolution de notre espèce, en particulier dans les derniers siècles ou dernières décennies (Cavalli-Sforza et al., 2003). La plupart des recherches à ce jour se sont attardées aux gènes nucléaires, considérant souvent que l'ADNmt évoluait de façon neutre (c.-à-d. par dérive génétique) et qu'il n'avait qu'une

fonction limitée dans le processus d'adaptation des organismes à divers environnements (Ballard et al., 1995; 2014). À ce jour, on recense beaucoup d'études faisant le constat que le génome mitochondrial est aussi soumis à la sélection naturelle chez une variété d'espèces incluant l'humain (Galtier et al., 2009).

Plusieurs hypothèses ont été émises quant à l'origine de la sélection sur l'ADN mitochondrial. Parmi celles-ci, l'idée voulant que le climat soit une force sélective importante est assez récurrente. Dans une revue exhaustive Garvin et al. (2015) relatent environ 20 études où la principale conclusion est que le climat aurait contribué à la diversité mitochondriale actuelle. Les espèces concernées par ces études sont l'homme, plusieurs types de poissons (harangs, saumons, scombridés, etc.), quelques primates, des mammifères comme le lièvre, la chèvre et l'ours ainsi que certains oiseaux et la drosophile. La plupart du temps, l'étude porte sur la séquence d'ADNmt entière, mais parfois seulement sur certains gènes (ND2, ND3, ND5 et le cytochrome b). Les méthodes employées pour détecter la sélection vont des tests de neutralité aux propriétés physicochimiques des acides aminés et l'activité enzymatique des protéines mitochondriales.

Toutefois, le soutien à l'hypothèse climatique demeure jusqu'à maintenant équivoque et les résultats empiriques ne convainquent pas tous les chercheurs, principalement parce que les résultats de certaines études n'ont pu être reproduits avec d'autres ensembles de données, soit différents, soit plus larges (Sun et al., 2007). À titre d'exemple, la première étude à apporter un soutien à l'hypothèse climatique (Mishmar et al., 2003) s'est faite avec l'analyse de 104 séquences d'ADNmt humain. En utilisant le ratio de mutations non-synonymes/synonymes, ces auteurs ont relevé des associations entre la variabilité dans les gènes ATP6 et Cytb en lien avec différentes zones climatiques (Afrique, Europe et Sibérie). Ils en ont conclu que la répartition géographique des mitotypes avait fort possiblement été modelée par la sélection naturelle.

Cependant, d'autres recherches ont conclu que la relation entre la variabilité mitochondriale et le climat n'était pas suffisamment importante pour tirer de telles

conclusions. Moilanen et al. (2003) ont comparé 837 séquences d'ADNmt humaines, comprenant 53 des 104 séquences analysées par Mishmar et al. (2003). Leur analyse des rapports dN/dS dans les 13 gènes codants a révélé des différences au sein des haplogroupes mitochondriaux africains, asiatiques et européens en plus de la variation déjà documentée. Ces résultats suggéraient que le climat n'est peut-être pas la seule cause possible de sélection étant donné qu'il y a autant de variation à l'intérieur d'un groupe vivant sous un même climat que celle détectée entre deux groupes vivants dans deux climats différents. Elson et al. (2004) ont aussi étudié la relation avec le climat, cette fois avec 560 séquences d'ADNmt humaines et en utilisant sensiblement les mêmes groupes ethniques (Européens, Africains, Asiatiques et autochtones d'Amérique) sans réussir à reproduire les résultats de Mishmar et al. (2003). Ils ont aussi fortement critiqué la méthode employée par ces derniers, dont celle de Nei-Gojobori (1986) pour l'estimation des ratios dN/dS qui donne trop de poids à certains sites. Une critique additionnelle a été émise sur les prémisses de l'étude et l'incapacité de l'approche à séparer les effets du climat d'autres effets potentiels.

L'ensemble de ces recherches utilisait le test de ratio dN/dS , une méthode qui a été développée à l'origine pour comparer les gènes d'espèces divergentes pour obtenir de l'information sur l'évolution d'un gène entre la version ancestrale et moderne. Il a maintenant été démontré que lors de son utilisation au sein d'une seule population, ce test peut s'avérer inapproprié pour tester l'hypothèse de sélection naturelle (Kryazhimskiy et al., 2008). Les différences entre deux espèces divergentes représentent des polymorphismes qui se sont fixés au cours du temps alors qu'au sein d'une même espèce, l'échelle de temps est beaucoup plus courte et ces différences représentent des polymorphismes transitoires qui ne sont pas forcément fixés. Lors de l'utilisation de ce test avec une seule espèce, il est important de considérer que l'échelle de temps entre la divergence de deux populations d'une même espèce est beaucoup plus courte que le temps entre la divergence de deux espèces.

À ce jour, une très grande partie des recherches liant l'ADNmt et l'adaptation au climat repose sur des corrélations entre des signaux de sélection dans le génome et

différents environnements climatiques ou encore, entre les fréquences d'haplotypes mitochondriaux et des gradients associés au climat comme la température ou l'altitude (Camus et al., 2017). Comme il a été discuté à la section 1.3, les tests moléculaires sont sensibles aux événements démographiques passés, de même que la distribution des fréquences d'haplotypes. Cela est probablement exacerbé chez l'humain puisque l'expansion géographique et démographique s'est faite en même temps que la sélection présumée sur l'ADNmt.

Par contre, l'étude de Balloux et al. (2009) a été la première à utiliser une autre approche pour estimer l'impact de la dynamique démographique et du climat sur la répartition des fréquences des haplotypes mitochondriaux chez l'humain. En utilisant un modèle de régression linéaire, ces auteurs ont mesuré la diversité mitochondriale de plusieurs populations humaines en fonction d'un gradient de température en contrôlant pour la localisation géographique. Leur analyse intégrait donc la distance entre ces populations et l'Afrique subsaharienne (lieu d'apparition d'*Homo sapiens*) comme indicateur de la réduction de la diversité attendue à la suite des suites des migrations successives et de la colonisation des diverses régions de la planète. Cette étude démontre que les populations vivant dans des environnements plus froids ont une moins grande diversité mitochondriale et que la différenciation génétique entre les paires de populations est corrélée avec la différence de température. Ce qui soutient l'hypothèse de sélection par le climat.

1.7 Problématique

L'une des principales difficultés se présentant dans l'étude de l'évolution par sélection naturelle est qu'il s'agit d'un processus qui s'étale dans le temps, souvent sur des très longues périodes pendant lesquelles la sélection peut fluctuer (Bell, 2010). Cependant, il existe peu de jeux de données longitudinaux ayant une bonne couverture temporelle sur des populations naturelles d'espèces longévives. De plus, l'information génétique disponible est souvent limitée aux individus encore en vie. Il devient alors difficile de mettre en lien le bagage génétique des individus des générations précédentes,

leurs phénotypes et leur fitness pour étudier la sélection et les variations de fréquences des gènes dans ces populations.

Bien que les études réalisées jusqu'à maintenant apportent un soutien non négligeable à l'hypothèse du climat, les résultats disponibles sur la sélection naturelle de l'ADNmt en lien avec le climat sont le reflet d'épisodes de sélection survenus il y a longtemps. Par conséquent, ils se basent nécessairement sur la comparaison de patrons au niveau populationnel (et non individuel), pouvant être biaisés par des phénomènes, comme la migration différentielle des lignées, difficiles à mesurer rétrospectivement. Ce soutien à l'hypothèse du climat est donc indirect. Un test direct ne peut se faire qu'en mesurant la sélection naturelle sur l'ADNmt à partir de données individuelles de mitotypes et de fitness, de surcroît sur une échelle multi-générationnelle. À ma connaissance, une telle approche n'a pas encore été appliquée dans l'étude empirique de populations naturelles, encore moins humaines. En ce qui concerne les effets de l'ADNmt sur le fitness en lien avec le climat, les meilleures indications proviennent de recherches expérimentales chez la drosophile à l'intérieur de cages de perturbations où certaines lignées mitochondriales sont associées à une meilleure récupération physiologique après un stress de froid (Ballard et al., 2007). Camus et al. (2017) ont aussi observé que les drosophiles présentaient une tolérance variable au froid et à la chaleur en fonction de leur mitotype.

1.8 Mécanismes possibles d'adaptation

L'une des premières hypothèses émises pour expliquer le mécanisme par lequel l'adaptation au climat pouvait passer par l'ADNmt a été l'existence d'un compromis évolutif entre la production d'ATP et de chaleur (Mishmar et al., 2003), c'est-à-dire que des mutations dans l'ADNmt pourraient conduire à une production d'ATP moins importante, mais augmenter la production de chaleur et vice versa. Des mutations dans les gènes ATP6 et ATP8 découpleraient davantage la production d'ATP et seraient favorables dans les climats plus froids, permettant d'augmenter le métabolisme de base et la survie des individus.

Il existe des études à propos de l'efficacité du couplage ETC-ATP qui ont été réalisées à partir d'hybrides cytoplasmiques contenant un même ADN nucléaire, mais différents ADNmt, représentant majoritairement différents haplogroupes. Selon Wallace (2015) la variation fonctionnelle, sur le plan métabolique, entre les haplogroupes, aurait largement contribué à la radiation adaptative des populations humaines vivant sous différents climats. Parmi les techniques employées pour mesurer l'efficacité du couplage, on compte l'analyse cinétique modulaire (Brand et al., 2011). Celle-ci divise le circuit de protons en trois modules. Le premier comprend les réactions impliquées dans la génération du gradient trans-membranaire. Le deuxième comprend les réactions impliquées dans la transformation l'ADP en ATP. Le troisième concerne les autres réactions qui utilisent le gradient, mais qui ne produisent pas d'ATP. Le courant de protons dans le système est mesuré et est divisé entre la fuite de H^+ et la production d'ATP. On utilise aussi le ratio entre l'ATP produit et l'oxygène consommé en présence d'une concentration d'ADP connue comme indicateur de la capacité oxydative (ATP/O). Une autre mesure couramment utilisée pour mesurer le couplage ETC-ATP est le ratio de contrôle respiratoire (RCR). Il s'agit du rapport entre le taux de consommation d'oxygène pendant la phosphorylation de l'ADP et le taux après épuisement de l'ADP (Brand et al., 2011).

Ces études ont donné des résultats parfois contradictoires. Certaines ont trouvé des différences fonctionnelles entre haplogroupes (Gómez-Durán et al., 2010; Larsen et al., 2014), alors que d'autres n'en ont trouvé aucune (Amo et al., 2007; Amo et al., 2008). Les liens de causalité entre des mutations de l'ADNmt et la capacité oxydative semblent difficiles à prouver. Le cas de la neuropathie héréditaire optique de Leber (LHON) est un bon exemple. Le lien d'association entre les mutations la pénétrance de la maladie est connu depuis longtemps, mais les recherches *in vitro* ont mené à des résultats mitigés (Howell, 1998). Dans la maladie de LHON, au moins trois mutations causent des changements d'acides aminés dans le complexe I de l'ETC. Le polymorphisme G11778A change une arginine pour une histidine, G3460A une alanine en une thréonine et T14484C une méthionine en valine dans les gènes ND4, ND1 et ND6 respectivement. Il suffit d'avoir une seule de ces mutations pour développer la maladie (la pénétrance est toutefois

très variable d'un porteur à l'autre). On croit que ces mutations altèrent l'intégrité du complexe I et réduisent son activité spécifique en modifiant son affinité pour ses substrats ou son interaction avec d'autres protéines. Le résultat étant une production d'énergie insuffisante et menant à la dégénérescence des cellules ganglionnaires de la rétine, cellules qui ont une forte demande en énergie provenant des mitochondries. Selon les études, les résultats en laboratoire ont démontré soit une réduction significative de l'activité du complexe I et de la respiration mitochondriale chez les porteurs de ces mutations, soit une baisse d'activité du complexe I non accompagnée d'une réduction de la respiration ou tout simplement aucune baisse d'activité significative (Howell, 1998).

Il faut mentionner que les conditions dans lesquelles sont exécutées les mesures ne sont pas nécessairement le reflet des conditions *in vivo*. L'effet de certains variants mitochondriaux semble dépendre du contexte nucléaire, des interactions entre ADNmt et ADN nucléaire ainsi que les tissus dans lesquels se trouve l'ADNmt. De plus, une étude récente a démontré que l'emploi du ratio ATP/O et le RCR peut conduire à des difficultés d'interprétation (Selman et al., 2018). En utilisant la truite comme animal modèle, les auteurs ont remarqué que le ratio ATP/O était plus bas chez les truites soumises à un jeûne comparativement à celles qui étaient nourries alors que le RCR lui était plus haut chez les truites à jeun que chez les truites nourries.

L'étude de Selman et al. (2018) met en évidence l'importance que peut avoir l'apport nutritionnel sur la capacité oxydative de l'ETC chez un organisme complexe. Les effets d'un débalancement des nutriments sont bien connus au niveau cellulaire : modification de la structure des mitochondries et de ses fonctions comme il a été mentionné à la section 1.5. Toutefois, il est difficile de généraliser ces données de cellules en cultures à l'adaptation à long terme. Lorsqu'ils ont émis leurs hypothèses sur l'adaptation au froid Mishmar et al. (2003) ont d'ailleurs mentionné que la diète pouvait être un facteur important dans ce processus. Dans une revue plus récente Ballard et al. (2015) discutent de la possibilité que le régime alimentaire, changeant avec les saisons, puisse influencer l'avantage sélectif des variants mitochondriaux dans les populations naturelles. Ces auteurs développent un argumentaire théorique sur les effets épigénétiques

que pourrait induire le changement de diète et ses liens avec les voies de signalisation mitochondriales. Quoique cet article reste spéculatif, des études montrent qu'un apport déséquilibré de macronutriments peut avoir un impact significatif sur l'activité mitochondriale.

On trouve que certains régimes alimentaires (riches en gras ou riches en sucre) entraînent une modification du nombre de copies de l'ADNmt et des fonctions oxydatives chez des animaux modèles comme le rat, la drosophile et la souris (Cormier et al., 2019; Dassanayaka et al., 2015; García-Ruiz et al., 2014). Il semblerait aussi que le ratio protéine-carbohydate (P:C) de l'apport alimentaire chez la drosophile soit un facteur déterminant dans l'expression des traits d'histoire de vie selon le mitotype qu'elles portent (Aw et al., 2017; Sun et al., 2014; Towarnicki et al., 2017). Une réduction de l'activité du complexe I a aussi été observée en présence d'un ratio P:C plus faible (Pichaud et al., 2013). Une consommation limitée de méthionine a aussi été associée à une augmentation des fonctions mitochondriales et d'une réduction du stress oxydatif chez le rat (Sanz et al., 2006). Certains chercheurs donnent même des recommandations alimentaires aux personnes atteintes de maladies mitochondriales (Rinninella et al., 2018).

1.9 La mortalité infantile comme mesure de la sélection mitochondriale

Les mécanismes d'adaptation décrits précédemment m'ont aidé à déterminer quel trait d'histoire de vie était le plus susceptible d'être influencé et par le climat et par les variants mitochondriaux dans les populations humaines. Basé sur ces théories, ainsi que pour des questions pratiques, ce choix s'est arrêté sur la survie, plus particulièrement la survie infantile.

Les nouveau-nés sont très vulnérables à l'environnement dans lequel ils se développent. Ils sont particulièrement sensibles au froid. L'hypothermie est un facteur de risque de morbidité et de mortalité des nouveau-nés selon l'Organisation mondiale de la santé (Mullany, 2010). Dans les pays en développement, de 32 à 85 % des enfants nés dans les hôpitaux subissent au moins un épisode d'hypothermie et ce taux se situe entre

11 et 92 % pour les enfants nés à la maison (Lunze et al., 2013). De façon surprenante, l'hypothermie plus présente dans les climats tempérés que dans les climats froids, principalement là où il y a de grands écarts de température entre le jour et la nuit. Chez les adultes, le frissonnement est un bon moyen de garder sa chaleur, alors que chez les bébés, en l'absence de contraction musculaire, la thermogenèse passe par le tissu adipeux brun (TAB). C'est entre autres, l'un des mécanismes reconnus d'adaptation au froid chez les petits mammifères (Cannon et al., 2004). Récemment, une recherche chez le campagnol des champs (*Microtus Arvalis*) a démontré que cette thermogenèse était significativement influencée par le génome mitochondrial ainsi que par des interactions mito-nucléaires (Lowe et al., 2018).

La quantité de tissus adipeux bruns est directement liée au poids à la naissance, et ce dernier est l'un des facteurs les plus déterminants pour la survie périnatale (Dunger et al., 2007). Une certaine quantité de tissu adipeux est donc nécessaire pour la croissance normale du nourrisson. Afin qu'un enfant naisse avec une quantité de TAB suffisante, il doit avoir une croissance fœtale optimale. Or, depuis longtemps, on reconnaît l'existence d'un mécanisme restreignant la croissance fœtale se transmettant par les lignées maternelles (Ounsted et al., 1986), attribué quelquefois à l'ADNmt (Casteels et al., 1999). La croissance fœtale est aussi intimement liée au statut nutritionnel de la mère. La taille du bébé à la naissance, ainsi que la taille du placenta ont été associées au taux de glucose sanguin maternel (Dunger et al., 2007), à la balance entre les protéines et les carbohydrates et à la composition en protéine de la diète maternelle (Moore et al., 2004). Il a aussi été démontré que les dysfonctions mitochondriales exacerbent l'effet d'une diète à faible teneur en protéines sur la réduction du poids néonatal et entravent l'homéostasie du glucose après la naissance (Malik et al., 2017). Un autre rôle possible des variants mitochondriaux dans la restriction de la croissance fœtale serait attribuable à la production de ROS. Ceux-ci joueraient un rôle dans le développement de la prééclampsie et des fausses couches (Díaz et al., 2014; Duhig et al., 2016). Cette production de ROS peut possiblement être influencée par des mutations dans l'ADNmt et par le régime alimentaire (Cardoso et al., 2013; Vives-Bauza et al., 2006).

De plus, plusieurs maladies mitochondriales se déclarent très tôt dans l'enfance. Le Tableau 1.1 présente quelques cas parmi les plus étudiés. Ces problèmes associés à l'ADNmt peuvent augmenter la mortalité. En conséquence, la survie infantile est une composante du fitness qui contribue vraisemblablement de façon importante à la sélection sur les variantes de l'ADNmt. Il semble aussi que la disponibilité et la composition des ressources nutritionnelles pouvant varier d'une région à l'autre, varier de façon saisonnière ou selon le climat local (pluie abondante, sécheresse, hiver long) puisse imposer une contrainte sélective importante la variation sur l'ADNmt par le biais de la survie infantile.

Tableau 1.1

Pathologies mitochondriales transmises maternellement
et s'exprimant parfois dans l'enfance

Syndrome	Phénotype(s)	Mutations
Neuropathie optique de Leber	Neuropathie optique	3460 G > A 11778 G > A 14484 T > C
Syndrome de Leigh	Retard psychomoteur, dystonie, convulsions, mouvements oculaires anormaux, vomissements récurrents, anomalies respiratoires	8993 T > G + 26 autres
Faiblesse musculaire neurogène	Neuropathie, ataxie, rétinite pigmentaire	8993 T > G 8993 T > C
Épilepsie myoclonique avec fibres rouges déchiquetées	Myoclonie, épilepsie, myopathie	8344 A > G 8356 T > C
Encéphalomyopathie mitochondriale, acidose lactique, pseudo-épisodes vasculaires cérébraux	Myopathie, acidose lactique, épisodes ressemblant à un AVC, encéphalopathie, convulsions	3243 A > G 3271 T > C

Ce tableau a été reconstruit à partir de Saneto et al. (2013).

1.10 La mortalité infantile

Le taux mortalité infantile est classiquement défini comme la proportion d'enfants nés vivants et qui décèdent avant d'atteindre l'âge d'un an. Il est commun de la séparer en deux causes, soit celles endogènes, c'est-à-dire résultant du développement fœtal, et exogènes, c.-à-d. dépendantes du milieu de vie post-natal. Cette séparation peut se faire facilement lorsque la cause du décès est connue. Plus généralement, les causes de décès ne sont pas disponibles dans les données historiques et lorsque l'on veut séparer ces deux types de mortalité infantile, on considère que les décès survenus avant le premier mois sont de cause endogène car ceux-ci surviennent dans les premiers jours de vie (Bourgeois-Pichat, 1951). Dans le cas de cette étude, la précision des données historiques ne nous permet pas nécessairement de travailler à la journée près. C'est pourquoi aucune division de la mortalité infantile n'a été faite ici. Les informations sur les dates de décès des jeunes enfants contenues dans les registres paroissiaux dépendent fortement de la déclaration par les parents ou correspondent au moment de la rédaction de l'acte de décès par le curé. Ces registres étant tenus par l'Église catholique, le niveau de croyance et de pratique des parents influence leurs déclarations (Levine, 1976). Les Canadiens- français ayant été de fervents catholiques à cette époque, ce problème s'avère toutefois moins important dans leur cas.

Dans les populations préindustrielles, la mortalité infantile était communément élevée. Comme l'indique le Tableau 1.2, les taux pouvaient fréquemment atteindre de 20 à 30 % (Bergeron et al., 2014; Sorg et al., 1983). Dans la population canadienne-française, la mortalité infantile a fait objet de recherches approfondies, en particulier dans la thèse doctorale de Amorevieta-Gentil (2009). Elle a identifié plusieurs facteurs de risque expliquant le taux élevé de mortalité infantile préindustrielle. Ainsi, l'âge de la mère à la naissance de l'enfant, le rang de naissance, la taille de la famille, de même que les intervalles protogénésique (le temps entre le mariage et la première naissance) et intergénésiq (le temps entre deux naissances successives) influencent considérablement les probabilités de décès des nourrissons. Plusieurs de ces variables sont toutefois corrélées entre elles, par exemple, l'âge de la mère et le rang de naissance, ou encore le

rang de naissance et la taille de la famille, ce qui peut rendre difficile la séparation de l'influence spécifique de chacune.

L'âge optimum d'une femme pour avoir des enfants se situait entre 25 et 30 ans (tranche d'âge où la mortalité infantile était la plus basse). Avant l'âge de 25 ans, on croit que le grand taux de mortalité infantile peut s'expliquer par des soins parentaux de moindre qualité en raison de l'inexpérience de la mère, par un développement maternel incomplet ou un risque accru de mortalité maternelle lors du premier accouchement. Après l'âge de 30 ans, des facteurs physiologiques pouvaient mettre en péril la survie des enfants. La maternité tardive est en effet caractérisée par une plus grande probabilité de complications obstétricales, de problèmes fœtaux et périnataux (Gillespie et al., 2013; Tarín et al., 1998). L'absence de moyens de contraception pouvait aussi conduire à des intervalles courts entre les naissances, augmentant les risques de complications des grossesses. Un rang de naissance élevé pouvait réduire la disponibilité, pour le nouveau venu, de ressources énergétiques devant être partagées entre plusieurs frères et sœurs, et ainsi augmenter les probabilités de décès. Finalement, les grossesses gémellaires étaient aussi un facteur de risque.

Tableau 1.2

Distribution de la mortalité infantile dans différentes populations préindustrielles

Population	Dates	Rate per 1000	Source
French Canadians	1700–1750	246	Henripin & Péron, 1972
Andover	1700–1759	152–156	Greven, 1970
France	1740	280	6 ^e Rapport, 1977
	1840	170	
Finland	1776–1800	210	Turpeinen, 1979
N. Shropshire, England	1750–1800	130–170	Jones, 1976
Madawaska French	1791–1838	132	

Tableau provenant de Sorg et al. (1983).

1.11 Objectif de la recherche

L'objectif général de cette recherche est de faire le premier test direct du rôle de la variation dans l'ADNmt dans l'adaptation au climat chez l'humain. Pour ce faire, j'utilise des données individuelles d'haplotypes et de fitness dans une population humaine depuis sa fondation afin de quantifier la sélection naturelle qu'impose le climat sur différents variants mitochondriaux et son effet sur l'évolution de leurs fréquences. Il s'agit de la population canadienne-française du Québec, plus précisément celle du Saguenay-Lac-Saint-Jean. Grâce au registre de population du projet BALSAC, nous connaissons les naissances, mariages et décès survenus dans cette population à l'ère préindustrielle, de même que la généalogie de ses habitants. Il est ainsi possible de reconstruire leur histoire de vie individuelle (survie, mariage, fécondité, longévité). De plus, étant donné le mode de transmission maternel de l'ADNmt, chaque haplotype connaît une histoire généalogique unique, le long de lignées maternelles, chacune partant d'une tête de lignée ou « fondatrice ». Cela nous permet de jumeler des données moléculaires avec les données généalogiques. Nous sommes donc en mesure d'avoir des données individuelles d'haplotypes mitochondriaux liées à des traits d'histoires de vie individuels pour une très grande partie de cette population, approchant les 80% entre 1840 et 1910. Les données moléculaires proviennent de l'Échantillon de référence québécois en épidémiologie génétique et génétique des populations du Québec. Cet échantillon est formé d'individus originaires des diverses régions du Québec qui sont représentatifs de la population en général, sans lien cibler de maladies particulières (portail web : <http://www.quebecgenpop.ca>).

1.12 Approche expérimentale

Nous assignons des haplotypes mitochondriaux échantillonnés sur des personnes vivantes aujourd'hui à des individus ayant vécu dans le passé. Pour y parvenir, nous utilisons une méthode d'imputation probabiliste adaptée de la méthode de Doyon et al. (soumis), voir section 2.3. Ces données moléculaires proviennent du laboratoire de Damian Labuda au CHU Sainte-Justine, du laboratoire de Catherine Laprise de

l'Université du Québec à Chicoutimi, du laboratoire de Bernard Brais de l'Université McGill ainsi que du laboratoire d'Emmanuel Milot, de l'Université du Québec à Trois-Rivières. Ces données ont servi pour des études sur la diversité mitochondriale au Québec (Moreau et al., 2011; 2009). La région de contrôle était déjà séquencée par la méthode Sanger et l'haplogroupe identifié avec des enzymes de restriction.

Nous avons centré notre étude sur la population du Saguenay-Lac-Saint-Jean, plus particulièrement sur les individus qui y sont nés entre 1842 et 1910. En effet, il s'agit de la région et la période de temps pour laquelle les données individuelles sont les plus abondantes. C'est aussi la période pour laquelle la couverture mitochondriale imputée était la plus grande. Cette couverture correspond à la proportion de la population historique pour laquelle un haplotype mitochondrial a pu être imputé. Nous avons utilisé des reconstructions climatiques annuelles couvrant toute cette période pour connaître les fluctuations climatiques vécues par les individus de cette population. Chacune de ces variables est décrite dans la section 2.3. Comme il a été décrit précédemment, nous mesurons la sélection s'exerçant sur l'ADNmt en utilisant la mortalité infantile comme indicateur de fitness. Il est en effet connu que la survie infantile, surtout à l'ère préindustrielle, pouvait être affectée par le froid et la disponibilité des ressources nutritionnelles. Nous avons donc utilisé un indicateur de la rigueur de l'hiver, soit l'Oscillation nord-atlantique (NAO), un indicateur de la productivité végétale, soit l'indice d'irradiance solaire (SI) et l'indicateur de sécheresse de Palmer (PDSI).

La mesure de la sélection naturelle sur les variants mitochondriaux ainsi que du rôle des variables climatiques sur cette sélection a été faite avec des modèles linéaires mixtes généralisés (GLMMs), implémentés dans le logiciel R. Un modèle mixte est un outil statistique permettant de modéliser la variation d'une variable dépendante à partir de prédicteurs (variables indépendantes). C'est une extension des régressions standard permettant d'intégrer à la fois des effets fixes et aléatoires afin de tenir compte de multiples sources de variabilité, de modéliser des distributions statistiques autres que la distribution normale ainsi que des matrices de covariance complexes.

Dans le but d'isoler au maximum la variation de la survie attribuable aux différents mitotype mitochondriaux, nous avons intégré dans les GLMMs des variables connues pour avoir un impact sur la survie infantile, et décrites à la section 1.10 (âge de la mère, rang de naissance, sexe, etc.). De plus, nous avons estimé la variation dans la survie infantile due à l'environnement familial partagé et avons pris en compte l'autocorrélation spatiale. En effet, lors de la colonisation du Saguenay-Lac-Saint-Jean, beaucoup d'individus ont vécu sur la vague du front de colonisation (front pionnier) et les mitotypes ont pu se répartir de façon non-aléatoire sur le territoire, en suivant le mouvement des familles. De plus, Moreau et al. (2011) montrent que dans cette population, le fait de vivre au front pionnier était associé à l'avantage sélectif représenté par les familles les plus fécondes, et par le fait même, au risque de mortalité infantile (plus de naissances). Ne pas tenir compte de l'autocorrélation spatiale pourrait donc mener à la détection de liens trompeurs entre la mortalité et la variation mitochondriale.

La mortalité infantile a été modélisée comme une variable binomiale (0 : ayant survécu au-delà de 1 an, 1 : étant décédé avant d'atteindre l'âge de 1 an). Le mitotype a été inclus comme une variable explicative catégorielle. L'effet d'interaction entre climat et mitotype, sur la mortalité infantile, a été inclus dans le modèle. Les résultats obtenus pour les différents haplogroupes et haplotypes mitochondriaux sont présentés dans le chapitre II à la section 2.3. Ce chapitre est présenté sous forme de manuscrit d'article. Le second chapitre comprend aussi une description de la mortalité infantile dans la population du Saguenay-Lac-Saint-Jean entre 1842 et 1910, une description de la méthode d'imputation utilisée ainsi qu'une description des variables climatiques choisies. Nous montrons dans cet article, que deux haplogroupes mitochondriaux sont sous sélection dans cette population. L'haplogroupe U est sous sélection positive tandis que l'haplogroupe J est sous sélection négative. Nous montrons aussi que deux haplotypes de l'haplogroupe J associés à des taux de mortalité infantile très proches répondent de façon similaire aux indicateurs climatiques. Nous pensons que ces résultats peuvent être issus de SNPs particuliers. C'est pourquoi, des tests de style GWAS, utilisant des séquences complètes ont été réalisés. Ces résultats sont toutefois préliminaires et ne font pas partis du chapitre II mais sont présentés et discutés au chapitre III.

CHAPITRE II

MTDNA AND CLIMATE ADAPTATION: A DIRECT TEST OF SELECTION USING MOLECULAR AND GENEALOGICAL DATA

Le contenu de ce chapitre est présenté sous forme de manuscrit en cours de préparation pour une soumission dans un journal scientifique.

Yannick Noël^a, Damian Labuda^{b,c}, Claudia Moreau^b, Catherine Laprise^d,
Bernard Brais^e, Emmanuel Milot^a

^a Département de chimie, biochimie et physique, Université du Québec à Trois-Rivières

^b Centre de recherche du Centre hospitalier universitaire Sainte-Justine

^c Département de pédiatrie, Université de Montréal

^d Département des sciences fondamentales, Université du Québec à Chicoutimi

^e Département de Neurologie et Neurochirurgie et de Génétique Humaine, Université McGill

2.1 Contribution des auteurs

Je suis l'auteur principal de cet article. J'ai mis en place des méthodes utilisées, produit l'ensemble des résultats et rédigé l'article. Les échantillons d'ADN proviennent de Claudia Moreau et Damian Labuda du CHU Sainte-Justine et de Catherine Laprise de l'UQAC ainsi que de Bernard Brais et Claude Berher de l'Université McGill. Emmanuel Milot a agi comme directeur de recherche et initié ce projet de recherche. Il a participé à la conception des méthodes utilisées et à l'interprétation des résultats.

2.2 Résumé de l'article

Nous investiguons ici l'hypothèse voulant que la variation dans l'ADNmt joue un rôle dans l'adaptation au climat et que ce dernier aurait façonné la répartition des variants mitochondriaux dans le monde. Utilisant une nouvelle approche, nous combinons des données généalogiques et moléculaires pour mesurer la sélection sur les variants de l'ADNmt par son effet sur la mortalité infantile (MI), ainsi que le rôle du climat comme agent de cette sélection. En utilisant la population préindustrielle Canadienne-française du Saguenay-Lac-Saint-Jean, nous soutenons l'hypothèse que la sélection sur les variants de l'ADNmt est dépendante du climat et de la disponibilité des ressources.

Nous brosons aussi un portrait des facteurs influençant la mortalité infantile dans la population du Saguenay-Lac-Saint-Jean entre 1842 et 1910 afin de bien comprendre les éléments pouvant être confondus avec la sélection. La sélection sur les variants mitochondriaux a été mesurée, à l'aide de modèles linéaires généralisés à effets mixtes, sur 8 haplogroupes mitochondriaux à partir de données sur 62 052 individus. Sur les 8, 2 sont fortement corrélés à la mortalité infantile : l'haplogroupe U et l'haplogroupe J. Nous avons ensuite quantifié la sélection sur les différents haplotypes représentés dans ces haplogroupes.

De plus, nous avons évalué l'impact du climat sur la sélection des variants mitochondriaux en utilisant des séries chronologiques d'indicateurs climatiques de la sévérité de la température, de l'hiver et de la sécheresse. Cet impact a été mesuré sous la forme d'interactions dans les modèles linéaires. Nous avons mis au jour une forte relation entre la variation de la mortalité infantile et celle de l'ADNmt qui dépend de l'indice d'oscillation nord-atlantique (NAO), ainsi que de l'indice d'irradiance solaire (SI). L'effet du climat seul sur la mortalité infantile a aussi été quantifié. Nous avons trouvé de fortes relations entre le NAO, le SI et le l'indice de sécheresse Palmer (PDSI). Nous trouvons donc un effet indépendant du climat et des variants d'ADNmt sur la MI ainsi qu'un effet d'interaction.

2.3 Article complet en anglais: mtDNA and climate adaptation: a direct test of selection using molecular and genealogical data

Abstract

Climate is thought to be a major selection agent that has contributed to shape regional distributions of mtDNA in human populations, but this hypothesis is still debated. To date, investigative methods have provided indirect evidence for episodes of selection over long timescales based on haplotype frequencies across several populations. By its important role in metabolic processes, mtDNA is hypothesized to modulate fitness traits that vary with climate. Here we combine genealogical and molecular data to measure selection on mtDNA variation, through its effect on infant survival, as well as the role of climate in shaping selection, using whole population scale multi-generational data spanning 70 years. Using the preindustrial French-Canadian population, we provide evidence for climate-dependent selection on mtDNA variation. Using generalized linear mixed-effect models, we found a strong response in the variation of infant mortality, which was associated to mtDNA variation and the North Atlantic Oscillation Index (NAO), as well as with the solar irradiance index (SI). This study provides the first direct evidence supporting the hypothesis that climate has been an agent of natural selection on mtDNA variation.

Keywords: mtDNA, natural selection, climate, adaptation, infant mortality

Introduction

In recent years, studies have shown or suggested that natural selection acts on mitochondrial DNA in several species, including humans (Ballard et al., 2004; Garvin et al., 2015). Climate is thought to be a major selection agent that has contributed to shape regional distributions of mtDNA haplotypes (Balloux et al., 2009; Mishmar et al., 2003; Ruiz-Pesini et al., 2004; Torroni et al., 2001). For instance, by its important role in metabolic processes, variation in mtDNA could be associated with differential survival in distinct regions of the world exposed to dissimilar climatic conditions. Mechanistic hypotheses include trade-offs between ATP and heat production that would confer a selective advantage in colder climates, to variation in dietary macronutrients among regions, that would modulate the concentration of metabolites and induce epigenetic changes (Ballard et al., 2015; Gershoni et al., 2009).

Although the idea that mtDNA plays a role in climate adaptation is still debated, evidence that it affects fitness traits is accumulating. Mitochondrial variants have been associated with variation in longevity, survival and fecundity in animals (Ballard et al., 2007; Camus et al., 2015; Christie et al., 2004; Ross et al., 2014). In humans, mitochondrial haplogroups were found to be associated with ageing rate (Dato et al., 2004), sperm motility (Feng et al., 2013) and antibody levels (Raby et al., 2007). Also, point mutations in mtDNA can lead to a wide array of diseases, perhaps the most widely known being the Leber hereditary optic neuropathy (LHON), which causes a degeneration of the optic nerve (Chalmers et al., 1999).

Infant survival is one fitness component that may be particularly sensitive to variation in mitochondrial functions induced by mtDNA polymorphism, and expressed under diverse climatic conditions, especially in preindustrial populations where infant mortality could reach very high values (Sorg et al., 1983). For instance, our data show that infant mortality (0-1-year-old) accounts for up to 59% of survival to reproduction in preindustrial Québec. Single mtDNA SNPs can be associated with variation in fitness components. One striking example is a variant (T14484C) causing LHON, which has been found to strongly reduce infant male survival in the preindustrial French-Canadian

population of Québec (Milot et al., 2017). The variant was also marginally associated with a lower probability of marriage and a lower fertility in male carriers surviving beyond 1 year old. If such associations between mtDNA variation and fitness (partly) reflected causal relationships, this would imply that some environmental factors exert a natural selection of mtDNA variants.

Effects of seasonal climatic variability on infant mortality are well documented in pre-industrial populations (Rocklov et al., 2014; Schumann et al., 2013). Newborns can be very sensitive to cold exposure (Derosas, 2009) and infectious diseases, a vulnerability that is influenced by weather (McKenney, 2001; Polgreen et al., 2018), and the nutritional environment, which can affect fetal growth, birth weight and neonatal survival (Lummaa, 2003; Malik et al., 2017; Moore et al., 2004). As size at birth is an important factor in child survival and that fetal growth restriction seems to be inherited through maternal line (Dunger et al., 2006), mitochondrial variants could be a cause of variability in infant mortality, particularly during gestation and in the first months of life. Climatic variation could exacerbate these effects, by shaping, for example, food abundance, disease outbreaks, and maternal health, thereby in utero conditions and parental care.

The methods used so far to detect climate-related mtDNA selection are mostly based on neutrality tests on genome data or empirical spatial distributions of mitochondrial sequence diversity (Camus et al., 2017). However, signatures left in DNA sequences or geographical distributions do not allow the direct measurement of selection, raising critiques of the climate hypothesis. Indeed, the mtDNA frequency spectrum observed across human populations is also compatible with expected effects of demographic processes, such as population size change, migration and divergence through random genetic drift or founder effects (Nielsen et al., 2009). The impact of these factors is hard to disentangle from that of natural selection (Ramírez-Soriano et al., 2008). Moreover, the functionality of a gene can only be fully assessed in the environment in which it is expressed. This kind of data is not available for signals of selection episodes that occurred over a long timescale. Here, we use a different approach to test “in real time” the climate hypothesis, focusing on infant mortality in a preindustrial French-Canadian population in Québec. By using genealogical imputation of mtDNA haplotypes (Doyon

et al., submitted), we can associate individual mtDNA and life history traits for people who were born in the decades following the foundation of this population in 1840. From these data we can measure selection on mitotypes generation-per-generation and its relationship with climatic variables, based on time series of climatic indicators of temperature, winter severity and drought. In this case, the use of the French-Canadian population is ideal because the foundation of New France, followed by the colonization of new territories that gave birth to subpopulations such as the Saguenay-Lac-Saint-Jean (SLSJ), exposed the newcomers to a large scale natural experiment: European genomes, adapted to a milder climate, were suddenly exposed to new conditions.

Methods

Data and study population

Genealogical, historical and demographic data available for the study period come from the BALSAC register (<http://balsac.uqac.ca>). This database contains deep and exhaustive pedigree information from marriage records collected by the Catholic church since the colonization of New France (early 17th century) until 1960 for the entire French-Canadian population of Québec. It also includes birth and death certificates for the Saguenay-Lac-Saint-Jean population. Colonization of this region began around 1840, with immigrants coming mainly from Charlevoix and the Côte-de-Beaupré counties, where rural areas were overpopulated (Hamel, 1993). Immigrants followed the spread of the timber industry in this part of the country, looking for new resources to supply the British market. This allowed the development of an agroforestry characterized by subsistence farming, while the forest provided extra resources and income for settlers (Séguin, 1976). The population expanded rapidly, starting with a population of 5,000 before 1852 and growing to 300,000 in 150 years (Gauvreau et al., 1988). Although there were immigrants from other Québec regions (30,000 between 1838 and 1911), after the initial wave of colonization the high natural fertility rate of the settlers was the main contributor to population growth (Robert et al., 2009). Individuals born in this region between 1842 and 1960 are known from their birth certificate, while those who died in

the same region by their death certificate. However, the data at our disposal contained information on births and deaths only up to 1917 for ethical privacy issues. Accurate data ended around 1910 though.

In the 1842-1910 period, about 73,000 children were born in 36 Saguenay-Lac-Saint-Jean parishes, 11,600 of whom died before the age of one year old, including reported stillbirths and interrupted pregnancies (Figure 1). Among them, 57% died within the first three month of life. Infant mortality was higher in male infants. It should be noted that these mortality rates are apparent rates since they do not control for under-registration of births and infant deaths that was more important on the colonization front, where church guidelines were not always strictly applied (Gagnon et al., 2009).

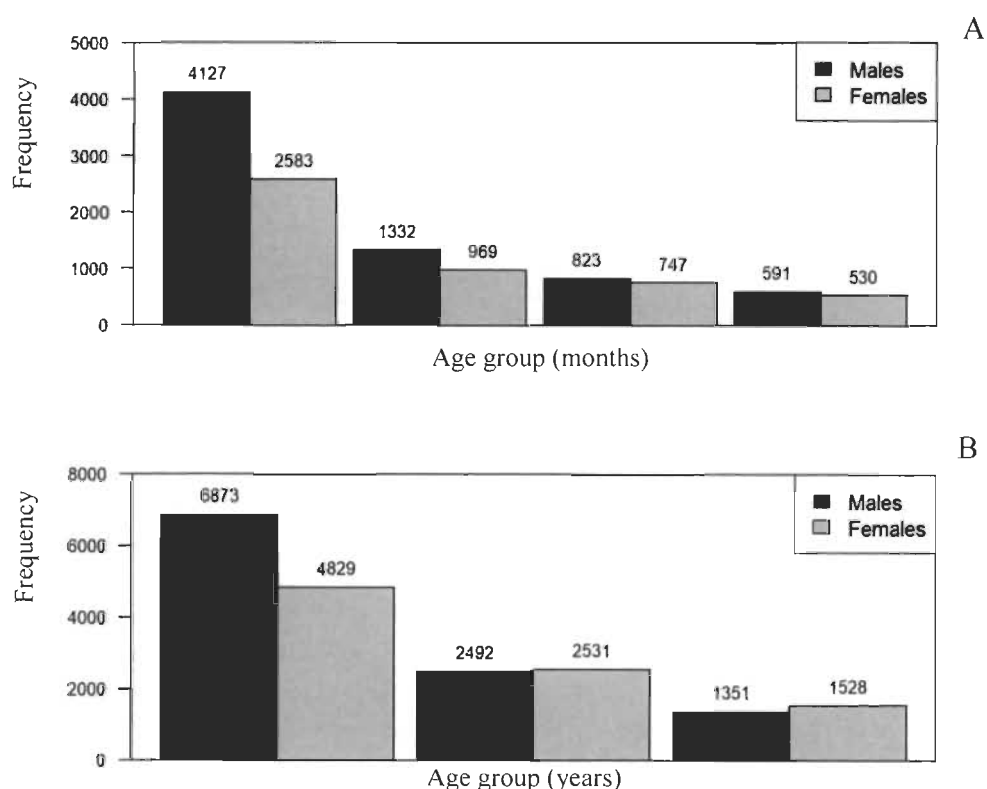


Figure 1 Distribution of deaths by age and sex in the Saguenay-Lac-Saint-Jean population.

Panel (A) shows the distribution of deaths from 0 to 1 year old in four age groups: 0 to 3 months, 3 to 6 months, 5 to 9 months and 9 to 12 months. Panel (B) shows the distribution of deaths from 0 to 16 years old in three age groups: 0 to 1, 0 to 5 and 5 to 16 years.

The infant mortality rate followed an upward temporal trend between 1840 and 1910, starting between 5 and 10%, rising near to 25% towards the end, with peaks around years 1847, 1866 and 1878 (Figure 2). This continuous increase in basal mortality rate, as the population grew, is possibly attributable to the development of urbanization and to an increase in the density of the population favoring infectious disease propagation (Gagnon et al., 2009).

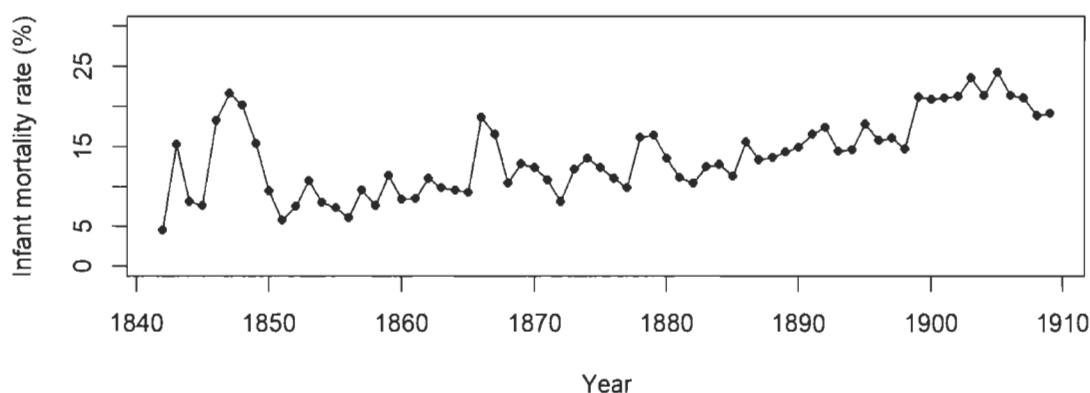


Figure 2 Infant mortality rate over the study period in the Saguenay-Lac-Saint-Jean population.

Mortality rate is calculated as the percentage of children born a given year who will die before the age of one year old. Some of them may have died the following year.

Highest mortality peaks typically correspond to epidemics and are identifiable using the death count of different age group: under 1 year, 1 to 5 year and 5 to 15 year (Bruckner et al., 2018; Figure 3). Common infectious diseases at that time were smallpox, typhus, cholera, dysentery, tuberculosis, influenza, intestinal parasites, measles, etc. (Landry et al., 2008). Mortality was also higher in spring and summer, which was likely caused by food shortage during months when the population still depended on the yield of the previous season. As noted by Landry et al. (2008), cold and food scarcity have rarely been a direct cause of death in preindustrial Québec but predisposed individuals to infection (Charbonneau, 1993). Spatial variation in infant death rates can also be observed according to the parish of birth (Figure 4).

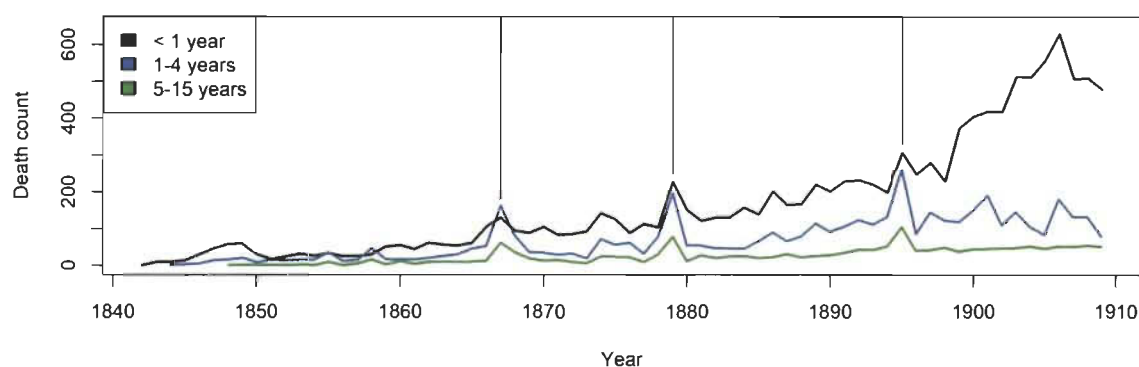


Figure 3 Death count by age group over the study period in the Saguenay-Lac-Saint-Jean population.

Trends in death count by year and age allow to note special events that suddenly increase mortality in all age groups (vertical lines, likely years of disease outbreaks). Age groups are divided as suggested by Bruckner et al. (2018).

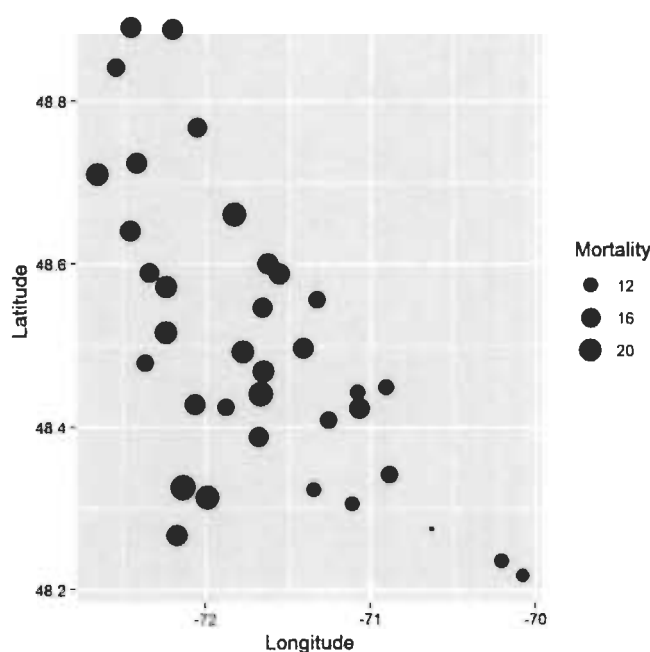


Figure 4 Infant mortality rate per parish for the whole study period in the Saguenay-Lac-Saint-Jean population.

Mortality rates are calculated by dividing the number of infant deaths by the number of births in each parish. Each bubble is a parish and its location on the graph is determined by geographical coordinates (in degrees). The size of the bubble increase with infant mortality rate (in percent).

Molecular data

All the mtDNA samples come from the Quebec Reference Group in Genetic Epidemiology and Population Genetics of Quebec. The sample is composed of 982 individuals for whom a saliva or peripheral blood sample was collected. Volunteers (all over 18) were recruited in the Montréal, Saguenay-Lac-Saint-Jean, Gaspésie, Côte-Nord, Québec, Abitibi-Témiscamingue, Lanaudière, Beauce and Outaouais regions. Part of these individuals (397) were selected if there was a possibility to build their genealogy up to at least two generations in the past and if they were not closely related, at least until the third generation (Moreau et al., 2009). All DNA samples were sequenced for region I (positions 15,998 to 16,400) and II (positions 30 to 407) of the hypervariable mtDNA control region (CR) by Sanger sequencing. Mitochondrial haplogroups were identified using restriction fragment length polymorphism (Moreau et al., 2011; 2011; 2009).

Probabilistic imputation of mtDNA

For each individual of a maternal lineage in the genealogy, the probability to have been a carrier of a given haplotype, among those sampled in the population, was calculated following the method of Doyon et al. (submitted). Briefly, genealogical imputation of mtDNA haplotypes was done with a probabilistic simulation approach accounting for potential mutations and genealogical errors. Genealogical errors include erroneous or missing pedigree links between parents and offspring. Common sources of errors include partial or wrong information, unreported adoption and extra-marital paternity. Random ceizures were generated in the pedigree, based on mutation and genealogical errors rates, dividing it into sub-pedigrees. Then, for each sub-pedigree, an haplotype was randomly chosen from the set of haplotypes carried by modern individuals connected to this sub-pedigree. This haplotype was imputed to the maternal founder of the sub-pedigree and to her descendents via the maternal line. This process was repeated to obtain at least 1,000 simulations compatible with the observed haplotypes in modern individuals i.e., simulations leading to one or more mismatches between observed and imputed haplotypes were discarded; see Doyon et al. (submitted) for details.

As their method was developed using BALSAC data, their genealogical error rate estimate of 0.36% per genealogical maternal link was applied here. The hypervariable region mutation rate of 0.43% per meiosis estimated by Sigurgardottir et al. (2000) was used. Simulations were done based on the control region SNPs. Thus, the 982 samples allowed to impute a control region haplotype to 62,581 of the 73,108 individuals in the Saguenay-Lac-Saint-Jean genealogy and a whole mtDNA sequences to 55,781 individuals, representing a coverage respectively of 85.6% and 76.3% of the data. Haplotypes from haplogroup H were by far the most common in this population with 22,790 carriers, followed by haplotypes from I and T haplogroups, carried by 10,042 and 9,885 individuals, respectively (Table 1). Among all mitochondrial haplotypes, most occurred at low frequency, with 21 of 158 having a frequency over 1% of the population. Haplotype diversity within haplogroup U was high compared to other haplogroups, with a total of 23 different strains found in 2,293 individuals.

Table 1. Mitochondrial haplogroups in the Saguenay-Lac-Saint-Jean population, number of haplotypes per haplogroup based on the control region SNPs compared to the Revised Cambridge Reference Sequence (Andrews et al., 1999).

Haplogroup	Number of carriers after imputation	Number of haplotypes per haplogroup	Number haplotypes with frequency $\geq 1\%$ *
A	215	3	-
H	22,790	69	9
I	10,042	6	1
J	3,017	10	2
K	8,461	12	4
N	183	1	-
T	9,885	19	3
U	2,293	23	-
V	4,330	6	1
W	1,244	4	1
X	121	5	-

* Number of unique mtDNA sequences (based on control region variants) that reach at least 1% in frequency in the Saguenay-Lac-Saint-Jean population.

Climatic variables

Climatic variables were chosen based on the availability of historical reconstructions and documented effects on organismal survival. Thus, we selected solar irradiance (SI) from Lean (2000), North Atlantic Oscillation Index (NAO) from Ortega et al. (2015), and the Palmer Drought Severity Index (PDSI) from Zhang et al. (2004). Some of these variables have been linked to infant mortality (Bergeron et al., 2014; Brown et al., 2016; Grosbois et al., 2005; Skjærvø et al., 2015), see Discussion.

Data for solar irradiance, NAO and PDSI are available from the National Climatic Data Center at <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets>. The NAO dataset was reconstructed from 48 time series records of ice cores, lake sediments, speleothems and tree rings. Values for winter (December to March) were used as an indicator of winter severity as NAO correlates with temperature, precipitation and storm tracks (Hurrell, 1995). Moreover, their effects on marine and terrestrial ecosystem are well established (Ottersen et al., 2001). In northeastern Canada, positive phases of the NAO are associated with colder winters (Wettstein et al., 2002). The SI dataset was reconstructed based on sunspot and faculae records and expressed in Watt/m^2 . It includes two time series, one with the annual mean irradiance of the 11-year solar cycle and the other one, with the 11-year cycle superimposed to a long-term background variability. SI was used here as an indicator of plant productivity (Currie et al., 1993; Pustilnik et al., 2004). The PDSI data is based on tree ring chronology and is available by grid point for North America, as well as a global measure for the continent. The point 269 was selected for our analyses, being the nearest one to the SLSJ region. PDSI standard values range from -4 to 4, from extreme drought to extremely wet weather respectively. Trends in these the climatic indices over the study period are shown in Figure 5.

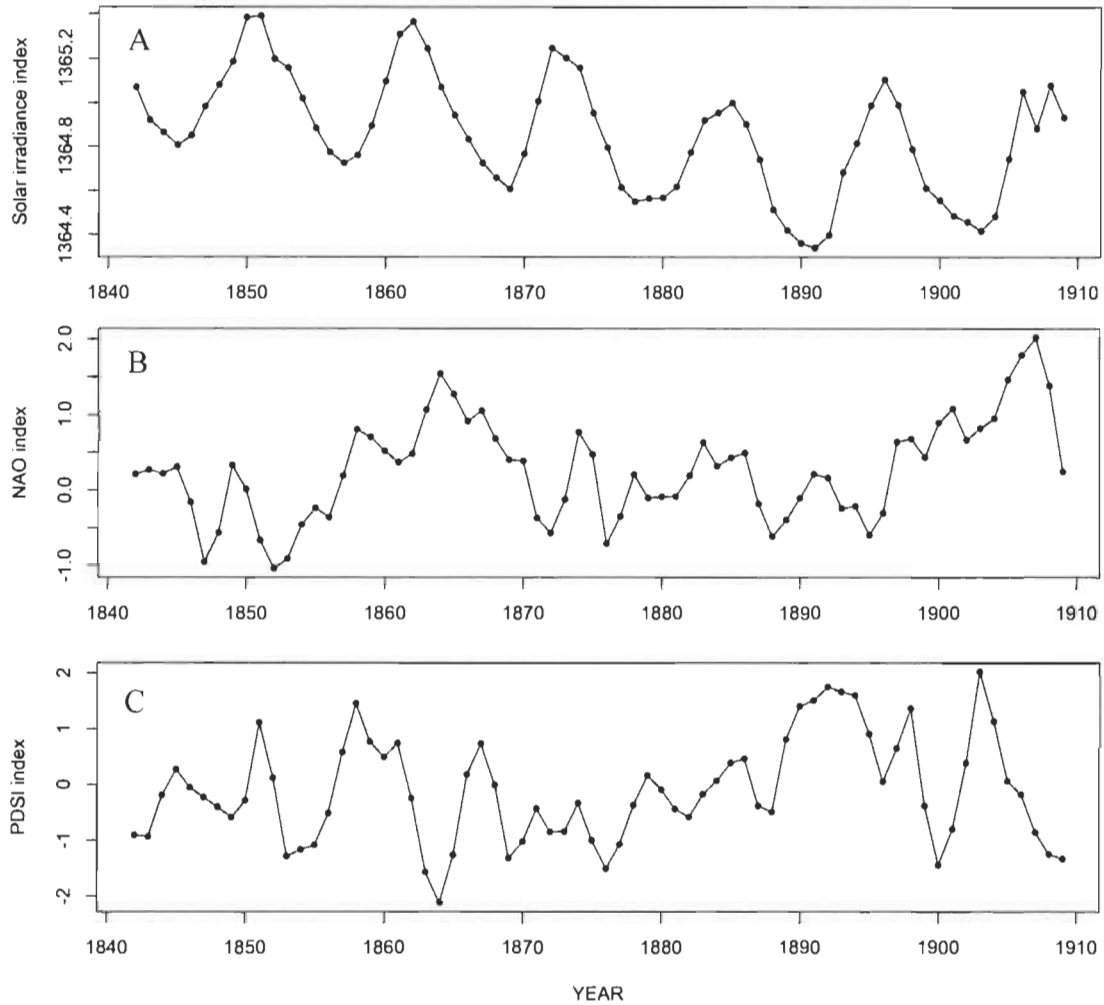


Figure 5 Time series for climatic indices over the study period (from 1842 to 1910).

A) The trend for the North Atlantic Oscillation. B) The trend for the solar irradiance. C) The trend of the Palmer Drought Severity Index.

Statistical analysis

In the assessment of infant mortality, every birth registered was considered. Individuals that were not related to a date of death were considered to have survived at least a year since they either left the region on their own as adults or left young with their family. We cannot totally exclude that in a few cases, an infant who left with his/her family died before age 1 outside the SLSJ region. The relationship between infant mortality, mtDNA, climatic variables and the mtDNA \times climate interaction were tested using generalized linear mixed-effect models with the spaMM package in R (Rousset et al.,

2014). Death probability was modeled as a binomial dependent variable (0: survived past one, 1: died before one). We controlled for fixed factors known or suspected to contribute to the variation in death probability. Fixed factors and interaction parameters to be kept in the final model were determined using likelihood ratio tests comparing models with different number of parameters. Fixed factors that were kept are: sex, birth year, twin births, birth season, birth rank, inbreeding coefficient of the mother. See Table 3 for model output. Birth rank and year were modelled as cubic, inbreeding as a linear covariate, and other variables as categorical. The age of the mother was not included in the model since it is lacking for the first 25 years of data and was highly correlated with birth rank for the remaining period ($r = 0.78$). The socio-economic status of individuals was not known, but as aforementioned, the majority of the population lived from subsistence farming and timber. The shared familial environment was also taken in account by fitting maternal ID as a random factor. Spatial autocorrelation was modeled by fitting a Matern covariance matrix with geographical coordinates of the birth parish.

We used the method of Brommer et al. (2008) to determine the time window to include in the analyses for climatic variables. For each variable, we selected the window for which the raw association with infant mortality was the highest. A 2-year window, i.e. up to two years before birth, was considered as making biological sense and thus, each time lag and all possible year-average combinations were tested. The best window for the NAO and the PDSI were the mean of the year of birth and the year before. So, if an individual was born in 1845, the NAO and PDSI value assigned to his birth year is the mean value of 1844 and 1845. For the SI, we chose the dataset that includes the background component. The best window was a two-year lag, which means that if an individual was born in 1845, the value of 1843 was used (Table 2).

Results

Climate effect on infant mortality

We found a positive relationship between infant mortality and the North Atlantic Oscillation, as well as with the Palmer Drought Severity Index (Figures 6 & 8) and a negative relationship with solar irradiance (Figure 7). Infant mortality increased linearly from 11.7% during years of low NAO values up to 15.4% in years of high NAO values (after correcting for fixed effects). The increase in IM follows a cubic trend as PDSI increases, from 9.6% during dry years to 12.6% during wet years. The cubic trend is most likely derived from the few extreme data points. The decrease in IM with respect to SI goes from 15.5% in years of low solar irradiance to 11.0% in years of high solar irradiance.

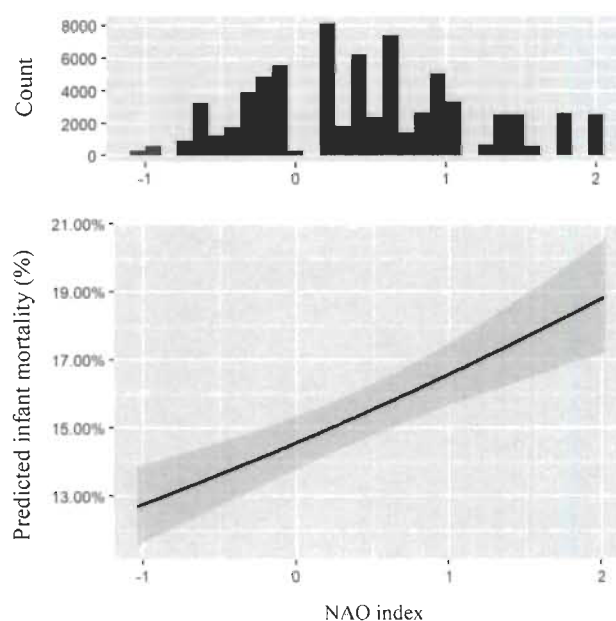


Figure 6 Relationship between the North Atlantic Oscillation and infant mortality with the distribution of each yearly value.

The Y axis is the probability of IM and X axis the NAO range observed in the SLSJ population. Straight line is the predicted effect obtain from GLMM and shading is the 95% confidence interval. The NAO is for the mean of the year of birth and the previous one.

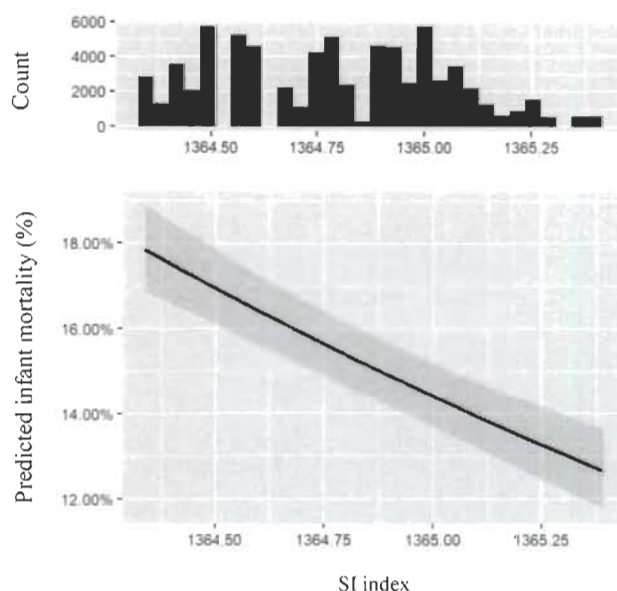


Figure 7 Relationship between solar irradiance and infant mortality with the distribution of each yearly value.

The Y axis is the probability of IM and X axis the SI range observed in the SLSJ population. Straight line is the predicted effect obtain from GLMM and shading is the 95% confidence interval. SI is for the second year preceding the birth year (i.e. two-year lag).

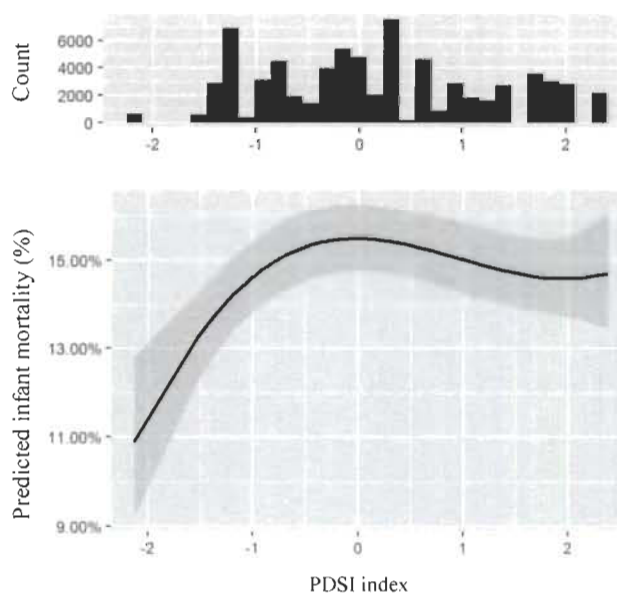


Figure 8 Relationship between the Palmer Drought Severity Index and infant mortality with the distribution of each yearly value.

The Y axis is the probability of IM and X axis the PDSI range observed in the SLSJ population. Straight line is the predicted effect obtain from GLMM and shading is the 95% confidence interval. PDSI is for the mean of the year of birth and the previous one.

Table 2. Glmm model output for the time window selection of climatic indices¹.

	Estimate	Standard Error	Z value	p value	Significativity	AIC	Choice
NAO	-0.0806	1.49E-02	-5.418	6.04E-08	***	61292	X
1 Year	-0.0626	1.67E-02	-3.742	1.83E-04	***	61308	
2 Year	-0.0116	1.95E-02	-0.597	5.51E-01		61322	
Mean 0-1	-0.1228	2.00E-02	-6.154	7.55E-10	***	61288	
Mean 0-2	-0.1556	2.68E-02	-5.799	6.65E-09	***	61291	
Solar	-0.0473	1.13E-02	-4.201	2.66E-05	***	61368	
1 Year	-0.0562	1.12E-02	-5.017	5.25E-07	***	61360	
2 Year	-0.0569	1.14E-02	-4.991	6.01E-07	***	61361	
Mean 0-1	-0.0725	1.14E-02	-6.339	2.32E-10	***	61345	
Mean 0-2	-0.4246	5.98E-02	-7.098	1.27E-12	***	61335	
Solar+BCK	-0.0318	1.21E-02	-2.618	8.86E-03	**	61379	X
1 Year	-0.0578	1.23E-02	-4.702	2.58E-06	***	61363	
2 Year	-0.0897	1.22E-02	-7.379	1.59E-13	***	61331	
Mean 0-1	-0.0487	1.25E-02	-3.907	9.35E-05	***	61370	
Mean 0-2	-0.0716	1.27E-02	-5.641	1.69E-08	***	61354	
PDSI	0.0224	1.17E-02	1.916	5.54E-02		61382	X
1 Year	0.0070	8.71E-03	0.808	4.19E-01		61385	
2 Year	0.0197	1.13E-02	1.745	8.10E-02		61383	
Mean 0-1	0.0272	1.21E-02	2.25	2.45E-02	*	61381	
Mean 0-2	0.0269	1.24E-02	2.165	3.04E-02	*	61381	

¹ “1 Year” and “2 Year” = lag of 1 or 2 years following birth year. “Mean 0-1” = mean value of the birth year and the previous one. “Mean 0-2” = value of the birth year and two years before birth. Estimate is the coefficient returned by the model and gives the magnitude of the effect, along with its standard error. Z value is the coefficient divided by the standard error (Wald statistic). The closer to zero the Z value, the more likely the population (“true”) coefficient of being zero. P-value is the probability of obtaining the observed results if the true coefficient is zero. *** p-value < 0.001, ** 0.01 ≥ p-value > 0.001, *p-value < 0.05. AIC is the Akaike Information Criterion and tells how well a model fits the data and allows avoiding overfit. Better models come with lowest values.

We found a significant effect of birth season on infant mortality. The risk of mortality was higher in spring than any other season, reaching 13.2% and at its lowest in the fall at 10.7%. (Figure 9). Moreover, there is a significant interaction between the NAO and birth season on infant mortality (Figure 10). Sorting seasons by decreasing order of the NAO effect on children gives spring, winter, summer and fall. Note that the NAO effect on IM decreases as the solar irradiance increases (Figure 11). Under harshest NAO, an increase of 1W/m^2 in SI decreases the death probability by about 4%.

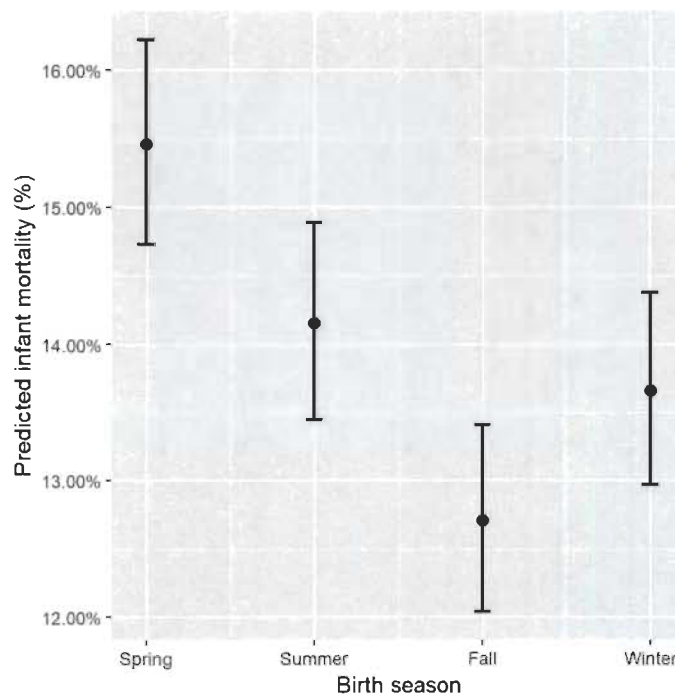


Figure 9 Predicted probability of IM by birth season (circles) and 95% confidence interval (vertical bars).

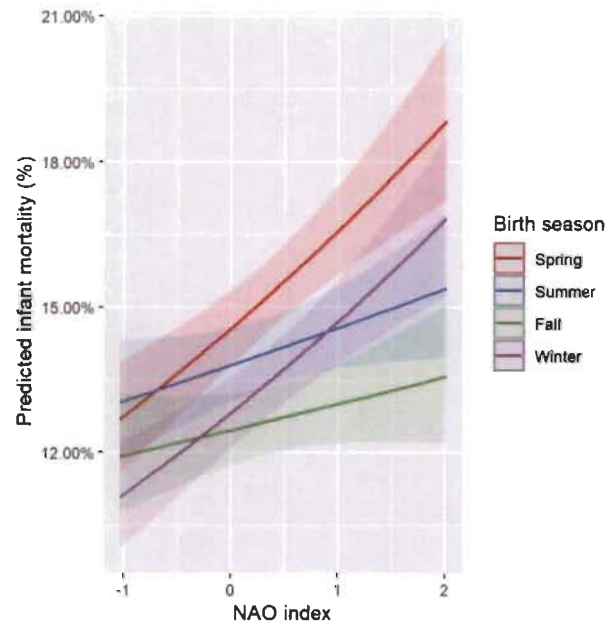


Figure 10 Relationship between NAO and infant mortality by birth season.
 The Y axis is the probability of IM and X axis is the NAO value as defined before. The solid line is the predicted effect of the NAO accompanied by its 95% confidence intervals.

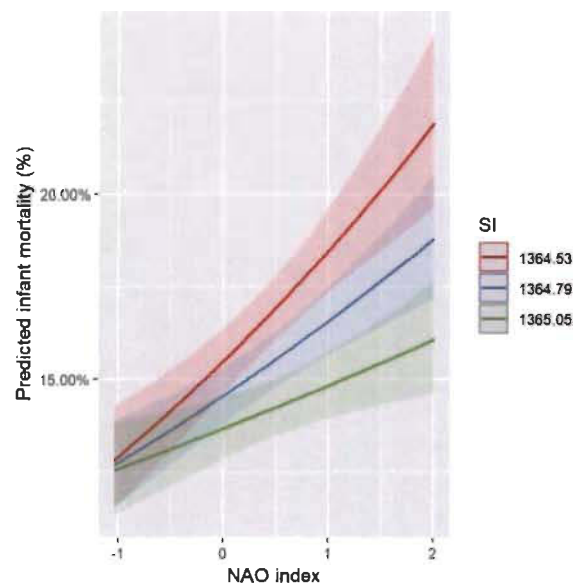


Figure 11 Relationship between NAO and solar irradiance.
 Y axis is the probability of IM and X axis is the NAO value as defined before. The solid line is the predicted effect of the NAO under three different SI accompanied by its 95% confidence intervals. Value of SI serve as graphical purpose. They illustrate the effect of min, middle and max value of the SI.

Table 3. Glmm model output for the effect of predictor on infant mortality.

Random effect			
	Estimate	Matern parameter	
Spatial autocorrelation	0.0685	nu = 0.082	rho = 0.038
Maternal effect	0.6531		
Fixed effect			
Predictor	Estimate	Std Error	Pr(> z)
(Intercept)	-1.570	0.0309	< 0.001
NAO	0.155	0.0304	< 0.001
SI	-0.051	0.0135	< 0.001
PDSI	-0.004	0.0204	0.834
PDSI ²	-0.058	0.0120	< 0.001
PDSI ³	0.024	0.0079	0.003
Female	-0.330	0.0209	< 0.001
Twin	1.715	0.0458	< 0.001
Birth rank	18.854	2.6342	< 0.001
Birthrank ²	33.101	2.6064	< 0.001
Birthrank ³	-13.022	2.5897	< 0.001
Birth year	62.389	3.577	< 0.001
Birth year ²	26.701	3.5569	< 0.001
Inbreeding of the mother	3.599	1.0623	< 0.001
Summer	-0.062	0.0355	0.080
Fall	-0.180	0.0368	< 0.001
Winter	-0.147	0.0363	< 0.001
NAO x SI	-0.0528	0.0164	0.001
NAO x Summer	-0.090	0.0391	0.021
NAO x Fall	-0.107	0.0411	0.009
NAO x Winter	0.004	0.0402	0.919

* rho: scalling factor for distance nu: smoothness

mtDNA haplogroups analysis

A first GLMM model included the mtDNA haplogroup as a categorical variable, along with climatic variables, but without an interaction between them. Number of individuals per mtDNA haplogroup are A: 215, H: 22,790, I: 10,042, J: 3,017, K: 8,461, N: 183, T: 9,885, U: 2,293, V: 4,330, W: 1,244, X: 121. Individuals with unknown haplogroups were excluded from the dataset ($n = 10,464$). After preliminary analyses, haplogroups A, N and X were also excluded because they represented a too small number of individuals ($n = 519$ altogether) to obtain reasonably precise estimates of their associated IM. The addition of mtDNA haplogroup in the model significantly improved the fit of the model (one-way Anova, $\text{Pr}(> \text{Chi}) = 2.7\text{e-}05$). Figure 12 shows predicted mortality for all haplogroups retained in the analysis. We detected positive selection (significantly higher survival) on mitochondrial haplogroup U and negative selection (significantly lower survival) on haplogroup J. All the others were within the same range with less than a percent of difference in IM.

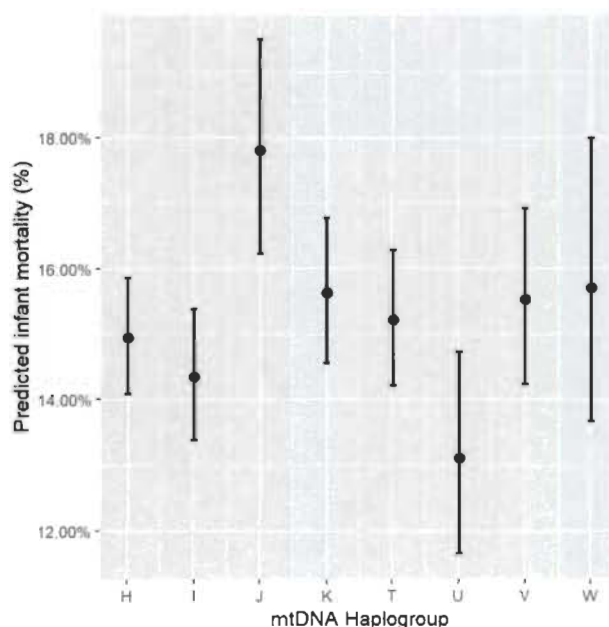


Figure 12 Probability of IM per haplogroup (circles) and 95% confidence interval (bars), as predicted by the glmm.

We then included the interactions between haplogroups and climatic variables. Three different model were generated, each with its specific interaction. We detected a weak significant interaction on IM between the NAO and haplogroup K, and a strong one between the NAO and haplogroups J and V (P -values respectively 0.065, 0.031 and 0.029; Figure 13A). IM for haplogroup J goes from 11.5% under low NAO values to 21.7% in years of high NAO. A similar positive interaction is observed with haplogroups V (IM going from 9.6% to 19.4%) and K (IM going from 10% to 18.4%). We found a strong negative interaction between SI and haplogroups J and U (Figure 13B). For haplogroup J, infant mortality decreases from 19.6% to 11.8% as solar irradiance increases from lowest to highest values. For haplogroup U, IM decreases from 15.3% to 7.4%. IM is lower for lower values of PDSI and peaks at intermediate PDSI (~ -0.5 to 0) for every mtDNA haplogroup, then decreases pass this point, except for haplogroups V and T, whose IM remains high (Figure 13C).

In accordance with the general pattern in Figure 12, haplogroup J is associated with a higher IM under most climatic scenarios, while the reverse is true for haplogroup U. Inter-haplogroup differences disappear under favorable NAO (Figure 13A). Haplogroup J is clearly disadvantaged when SI is low but shows a similar mortality rate to other haplogroups when SI is high. In contrast, mortality for haplogroup U is close to that of other haplogroups when SI is low and becomes noticeably lower when SI is high. Figure 14 shows the annual variation in IM caused by all three climatic variables predicted by the model for the whole study period. The trend for the whole population is illustrated, along with those for the two haplogroups found to be clearly under selection (U and J), which exhibit a strong variation with climatic gradient. Figure 15 shows that the distribution of births along climatic gradients is well balanced for the three indicators. Accordingly, haplogroup U increased in frequency in the SLSJ population by about 1% over the 75-year study period (Figure 16) and haplogroup J decreased by 1.2% during the same time. Frequency change for other haplogroups altogether were 0.005% from generation 1 to generation 4. It should be mentioned that data are less reliable for generation 4 since the date of birth of an individual were identified by the birth decade.

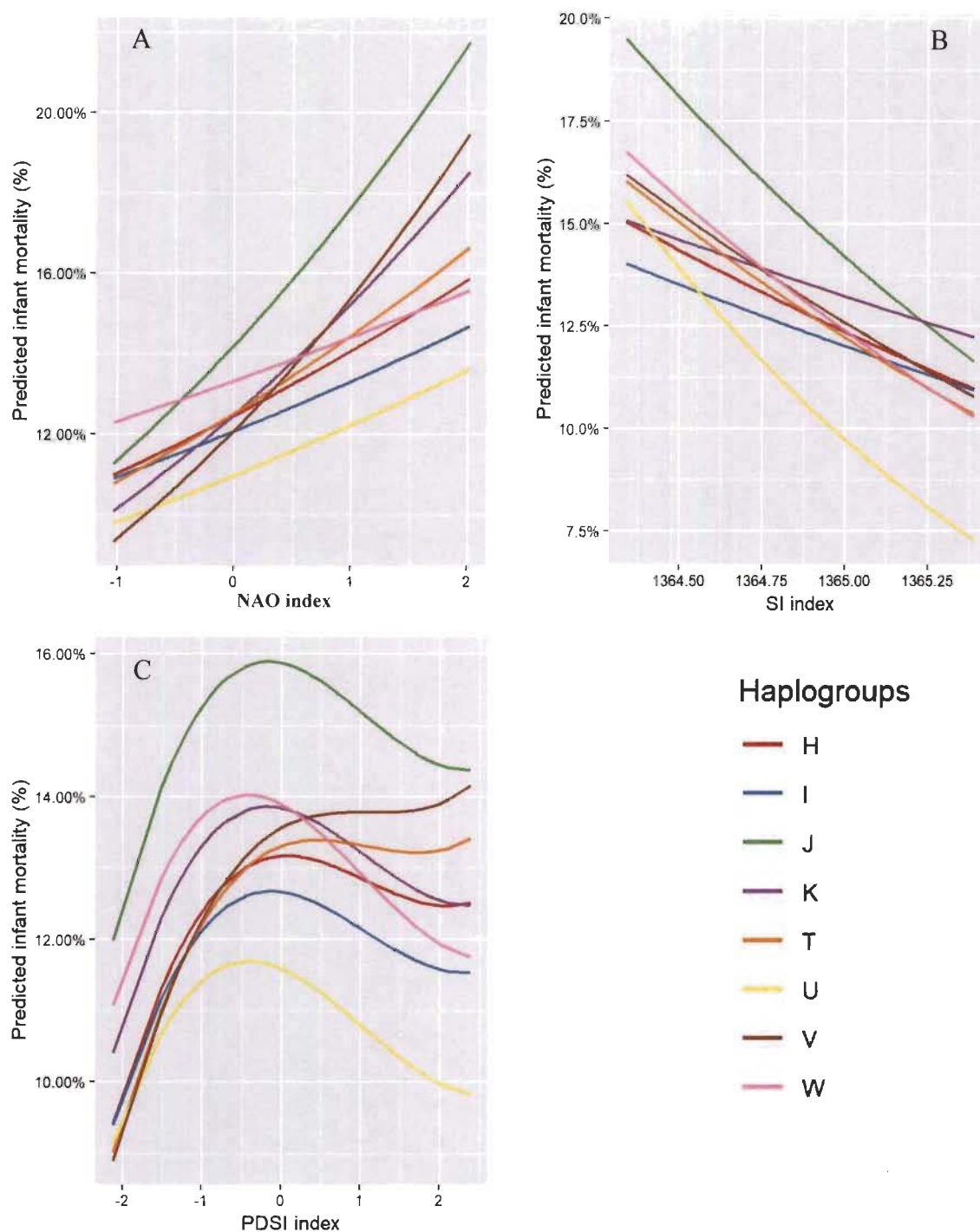


Figure 13 IM of mitochondrial haplogroups under three climatic gradients in the Saguenay-Lac-Saint-Jean population.

A) Interactions with the NAO. B) Interaction with the SI. C) interaction with the PDSI. Confidence intervals are not shown for clarity. For further information on each separate interaction, see supplementary material.

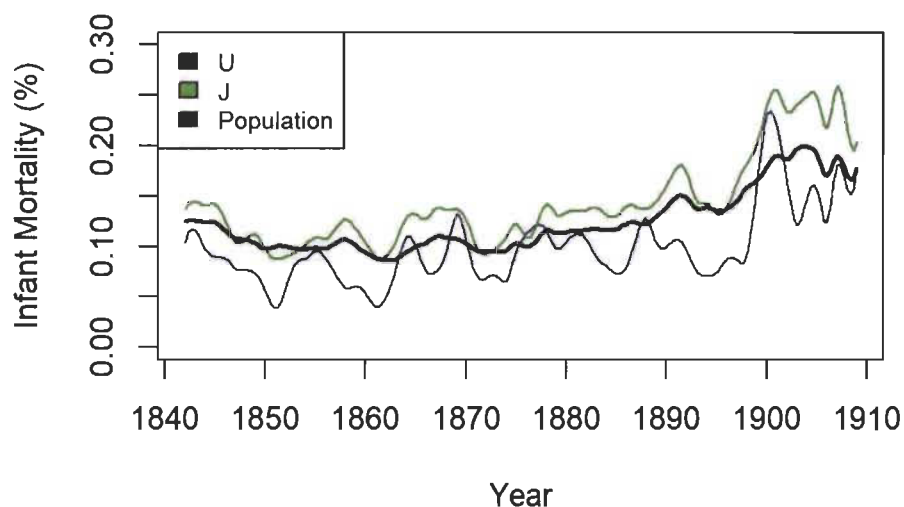


Figure 14 Year-by-year predicted IM probabilities for the entire study period in the Saguenay-Lac-Saint-Jean population.

Yearly IM fluctuations caused by the interaction of mtDNA haplogroups with all three climatic variables. Predictions are shown for haplogroup U (blue line), J (green line) along with the average for the whole population (black line). Only the year and climatic variables vary in this figure; others were kept constant

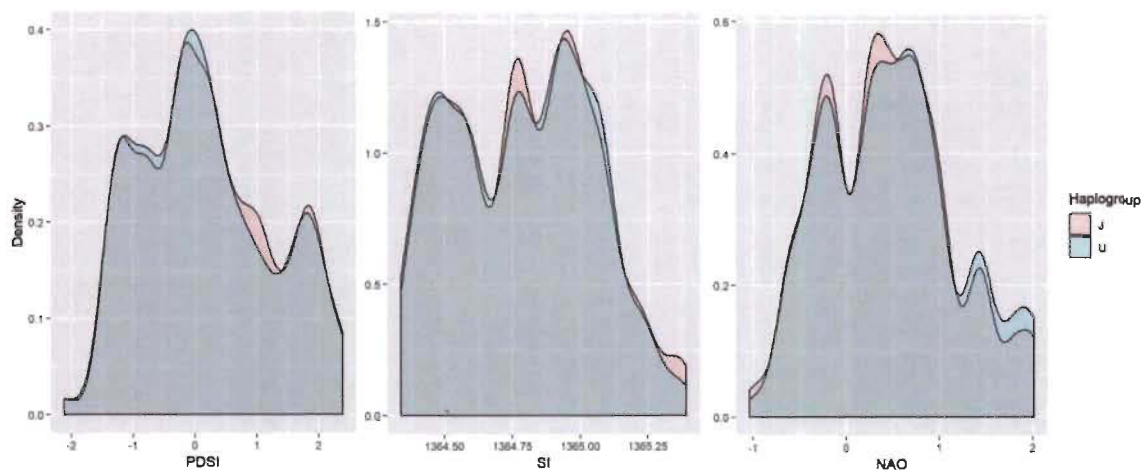


Figure 15 Distribution of the number of births among climatic gradient for the haplogroup U and J. From left to right: PDSI, SI and NAO.

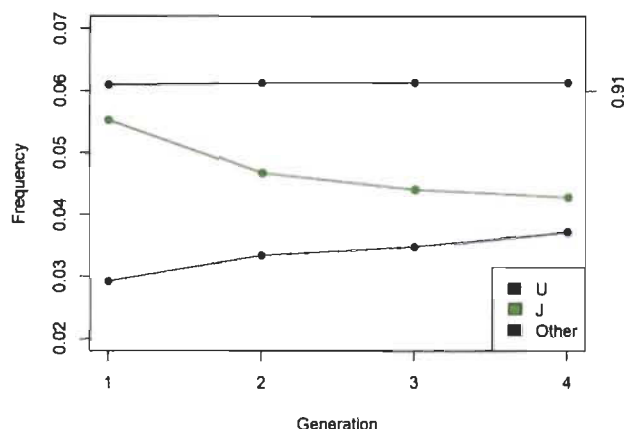


Figure 16 Temporal trends in the frequency of two mtDNA haplogroup under selection (U and J) compared to all other haplogroups pooled.

The frequency in the population was calculated after the removal of newcomers (immigrants) at each generation. Thus, for the three last generations, only the descendants of the first one was included in the analysis. Generation 1: 1840-1865, Generation 2: 1865-1890, Generation 3: 1890-1915, Generation 4: 1915-1940.

mtDNA haplotype analysis

After considering the results obtained from the haplogroup models, we assessed whether the effects observed in haplogroups U and J were caused by the haplogroup *per se* or driven by a subset of haplotypes belonging to these haplogroups. As before, we fitted models with and without interactions between climatic variables and mitotype. We filtered the dataset by keeping only individuals for whom the same haplotype was imputed for 90% of the simulations or more. Those below this 90% threshold were tagged as having an undetermined mitochondrial haplotype and pooled with other individuals. Such filtering was unnecessary in the haplogroup analyses because if a mutation had occurred somewhere in a lineage, it wouldn't change the haplogroup identity. Mitochondrial haplotypes were fitted as a categorical variable. Haplogroup J was split in two categories, namely CR haplotypes (HVS178_J, $n = 1633$, HVS283_J $n = 1044$). The other haplotypes from group J weren't pooled due to their rarity ($n = 144$ individuals in total). Haplogroup U was divided in two categories: HVS266_U ($n = 714$) and others grouped ($n = 1091$). Figure 17 shows IM predicted from the model for all groups retained in the analysis. We detect positive selection relative the pool of other haplotypes on one

mitochondrial CR variant from haplogroup U (HVS266_U) and negative selection on both CR variant from haplogroup J (HVS_178J & HVS_283J). We uncovered a strong interaction between the two J haplotypes and the NAO (Figure 18A). IM went from 10.9% under low NAO values for both J haplotypes, to 25.0% (HVS283_J) and 26.2% (HVS178_J) in years of high NAO. The slopes for both U groups did not differ much from that for the overall population. In the SI interaction model (Figure 18B), slopes are in concordance with effects reported above at the whole haplogroup level, except for the HVS266_U, which shows again a low IM and a gentle increase with SI, quite in contrast to the other groups. HVS283_J has showed a slight advantage (IM = 11.7%) over HVS178_J (IM = 13.1%) when SI is high. The decrease observed after the peak at -0.5 of the PDSI gradient among U and J haplogroup in the Figure 13C is only observable for the haplotype HVS283_J (Figure 18C). The other J haplotype remains stable, as well as the grouped U haplotypes. On the other hand, HVS266_U exhibits a completely different pattern with a continuous decrease in IM as PDSI increases (from 14.3% under drier years to 3.5% under wetter years).

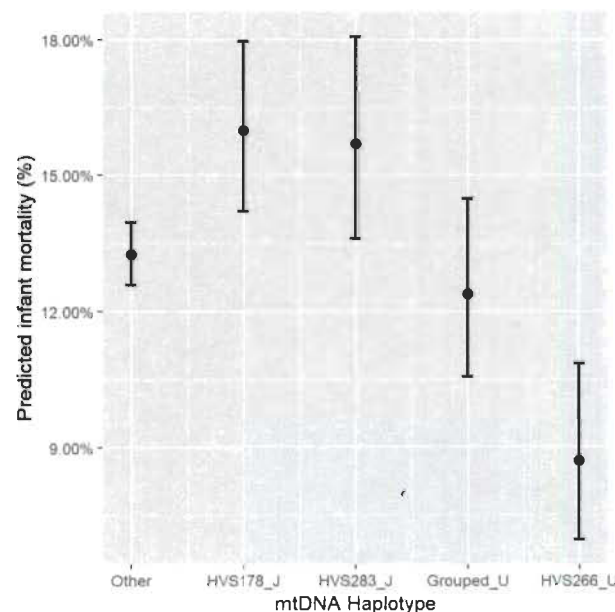


Figure 17 Predicted IM from the glmm model for mitochondrial haplotypes from U and J haplogroups.

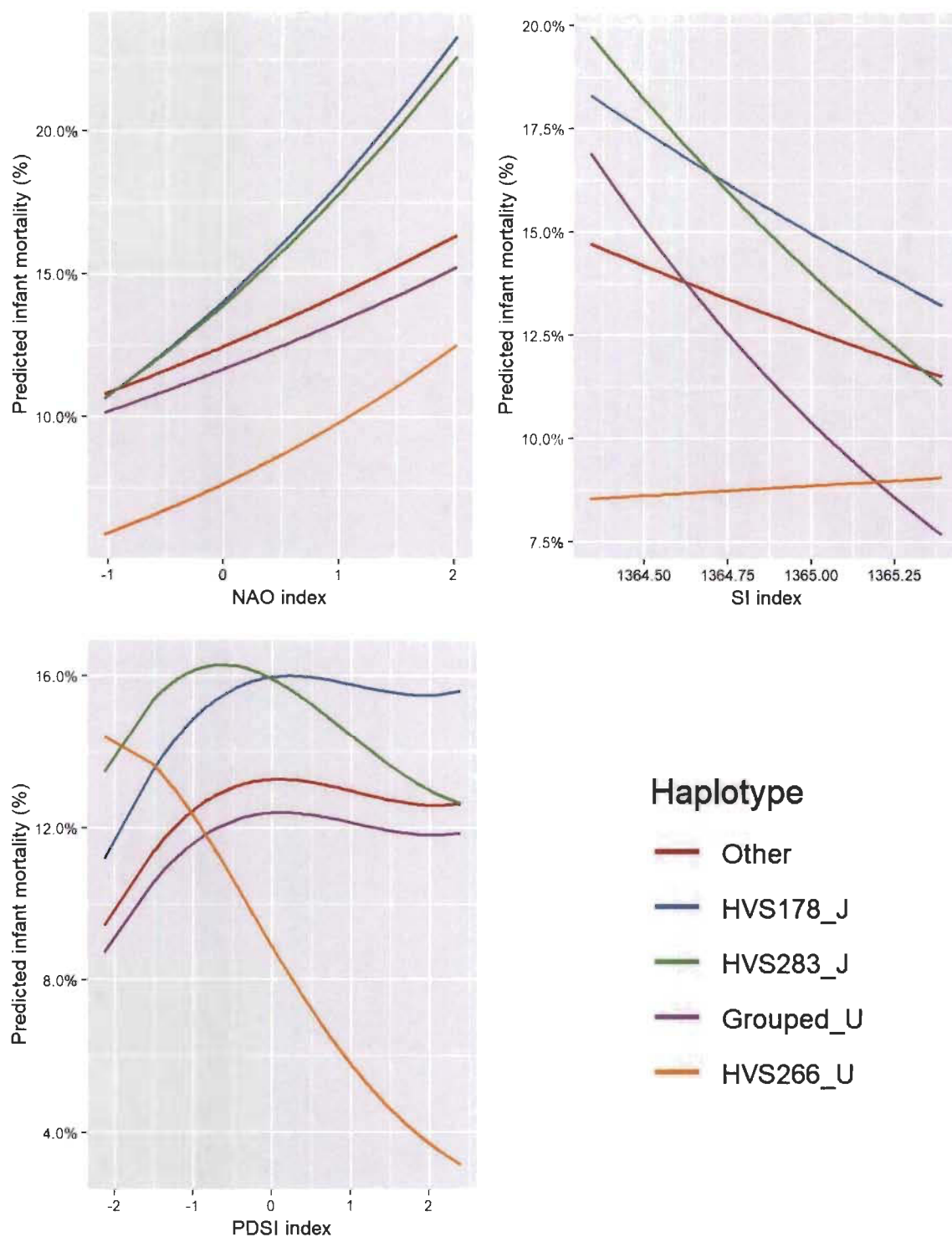


Figure 18 IM of mitochondrial haplotypes under three climatic gradients in the Saguenay-Lac-Saint-Jean population.

A) Interactions with the NAO. B) Interaction with the SI. C) interaction with the PDSI. Confidence intervals are not shown for clarity. For further information on each separate interaction, see supplementary material.

Discussion

We uncovered a clear relationship between climatic variables and infant mortality in the Saguenay-Lac-Saint-Jean population. This association was positive for North Atlantic Oscillations and Palmer Drought Severity Index, and negative for solar irradiance. Results for SI agree with the positive relationship found between this variable and infant survival at Isle-aux-Coudres, an islander subpopulation of the French-Canadian population (Bergeron et al., 2014). These authors found that infant survival probability increased with an augmentation of SI. On the other hand, they did not find any relationship with the NAO. However, their study was based on an older reconstitution of NAO made with less data, and thus may be less accurate. In addition, their sample was smaller (~5677 infants) due to the small size of the Isle-aux-Coudres population. Our findings for NAO agrees with others reported for mammals like Grosbois et al. (2005) who reported a linear decrease in female fulmar (*Fulmarus glacialis*) survival probability under high NAO value.

The population of the SLSJ shows a decrease in infant survival during harsher winters. The relationship we found between PDSI and IM is in contrast with that observed by Bruckner et al. (2018) in the French-Canadian population, using the Ancient Québec Population Register (before 1800). These authors observed a higher mortality during drought years in eastern Québec using an autoregressive model and data from precipitation and drought. Here, on the other hand, the lowest IM probability is under dryer years and the highest IM probability is under mild to wet PDSI value. However, while studying cliff swallows (*Petrochelidon pyrrhonota*), Brown et al. (2016) found that the selection coefficient associated with the PDSI were influenced by the colony size. Larger colonies were favored under wetter years and smaller one, in drier years, so there may be an effect of population density in the response to PDSI.

We also noted that IM was affected by seasonal environmental factors. Interestingly, we also found that the effect of NAO on IM depended on birth season. Furthermore, the NAO and SI effects on IM interact. The combined effect of the SI, NAO and birth season is not surprising if harsh winters and lower crop productivity lead to an increase of

IM (Scott et al., 1995) in other words, the negative impact of harsher winters on IM could be tempered when crop productivity is better. However, the association between NAO and birth season may also partly reflect the nature of the data and better capture variability in this time of the year. Variability in the NAO is larger from December to March than during the rest of the year. In addition, with the SI x NAO interaction effect and the highest death probability in spring, these results reinforce the hypothesis that malnutrition arises from food shortage after the winter. This situation may weaken the constitution of the mother during pregnancy which may lead to a higher risk of complication and lower birth weight of the newborn, making him more at risk under environmental stress. It is also thought that SI modulates NAO (Kodera, 2002) so that the former has a causal effect on the latter. The observed interaction with the NAO may come in part from this phenomenon. Our results for PDSI are more surprising since drought is generally associated with poorer crop yield (Leng et al., 2019). Hypothetically, the risk of contracting infectious diseases may be higher in wet years.

The improved fit obtained when adding mtDNA haplogroup in the model suggest that IM variance in the Saguenay-Lac-Saint-Jean is better explained by taking this parameter into account. This provides a direct support for the implication of mtDNA variation in IM. While the variance explained by mtDNA haplogroup may be partly confounded with maternal lineage effects (in particular X chromosome transmission and cultural transmission of behaviour), each mtDNA haplogroup was present in at least two different maternal lineages and we controlled for the shared maternal environment by adding a random maternal effect in the model.

Among all mtDNA haplogroups, J and U showed the strongest signals of selection with the lowest baseline IM of 10.9% for U and the highest one for J at 14.8%. These two groups also showed an impressive diversity of responses among different climatic scenarios. Haplogroup J had a significant difference in IM response for the NAO and SI gradient and U, a significant one with the SI. Moreover, IM probability stayed higher for J and lower for U under most values of NAO, SI and PDSI with non-overlapping confidence intervals under harsher climate. However, the addition of the interaction

between mtDNA haplogroup with any of the three climatic variables did not improved the fit of the model significantly but didn't reduce it either (reduction in AIC of 1 for the NAO, 8 for the SI and no change for the PDSI). This is not surprising considering that only three haplogroups (J, V and K) diverged from the population for the NAO interaction ($n = 15\,808$ of $62\,581$), about 25% of the population, two for the SI (J and U, $n = 5310$, ~8%) and none for the PDSI. Since a simpler model is preferable when AIC is similar, interactions between mtDNA haplogroup and climatic variable as a whole didn't explain as much variation as expected in IM in the Saguenay-Lac-Saint-Jean population but the NAO have a large impact on J, V and K haplogroup and the U and J are largely influenced by the SI. Another striking result was the higher IM variability in harsher (high NAO, low SI and high PDSI) than in mild climatic conditions. Thus, the differential selection linked to climate among mtDNA haplogroups is stronger under harsher conditions. We also found that fluctuations in climate are associated with important IM fluctuations over the entire period studied, in particular for haplogroups U and J (Figure 14). The distribution of births over the range of climatic values for haplogroup U and J are practically the same, suggesting that the difference in their response did not arise from heterogeneity in the data (Figure 15). In accordance with selection via IM, haplogroup J frequency decreased over 4 generations, and the pattern was reversed for haplogroup H. However, our analyses cannot totally rule out the possibility that those changes were due to random genetic drift. Nonetheless, a change in frequency of 1% over such short timescale is large and drift has a weaker influence when a population is growing (Slatkin et al., 1991).

We observed a similar pattern for two haplotypes of haplogroup J, in terms of both direct effects on IM and interactions with climate indicators, suggesting that the observed effects come from SNPs defining the haplogroup. On the other hand, it is less clear whether the selective advantage of haplogroup U comes from a single haplotype, because the HVS266_U haplotype was clearly advantageous compared to the other U haplotypes while the other U group also has a lower IM, but less pronounced. It had a different behavior with respect to SI and PDSI variation than what we uncovered for haplogroup U in general. Haplogroup U was found to be associated with a higher

immunoglobulin E level (Raby et al., 2007), lower ROS production, and a protective effect against some disease associated with ROS (Chen et al., 2012). Haplogroup J has been linked to lower ATP production, associated with type II diabetes (Feder et al., 2009; Wallace, 2015) and increased risk of developing the LHON disease for those who carry one of the causal mutations. While such associations are partly population specific (Dato et al., 2004), it is possible that SNPs defining haplogroup J create a genetic background that is more deleterious when associated with other mtDNA mutations, relative to other haplogroups found in the population. The advantage observed for haplogroup U can plausibly also come from the SNPs defining it, but with respect to the variation in IM within this haplogroup, it suggests that the beneficial effect may be tempered by other SNPs that may be present on different strains. Thus, SNPs defining haplogroup U may help reducing the deleterious effect of other mutations. Higher and lower fitness observed in J and U haplogroup respectively may also come from mito-nuclear interaction, whereby mutations on mitochondrial DNA modulate the effect of nuclear variants affecting mtDNA viability.

Conclusion

Our findings provide a direct support to the hypothesis that mitochondrial variation can play a role in adaptation to climate, here through its effect on infant mortality. Actually, our analyses suggest that the effect of mtDNA variation in IM is exacerbated by interactions with climatic factors. These associations also depend on the mitochondrial genetic background as well as mito-nuclear interactions, which were not modeled here. A next step would be to control for the nuclear genetic background by using the relatedness covariance matrix to quantify mito-nuclear epistasis. Another interesting option would be to get the full mtDNA sequence and see if some SNPs are common to mitochondrial haplotypes that have similar responses to climate gradients. Given the magnitude of the effects documented here at the scale of a local population, it is quite plausible that the climate contributed to shape the mitochondrial diversity in the world.

Supplementary material

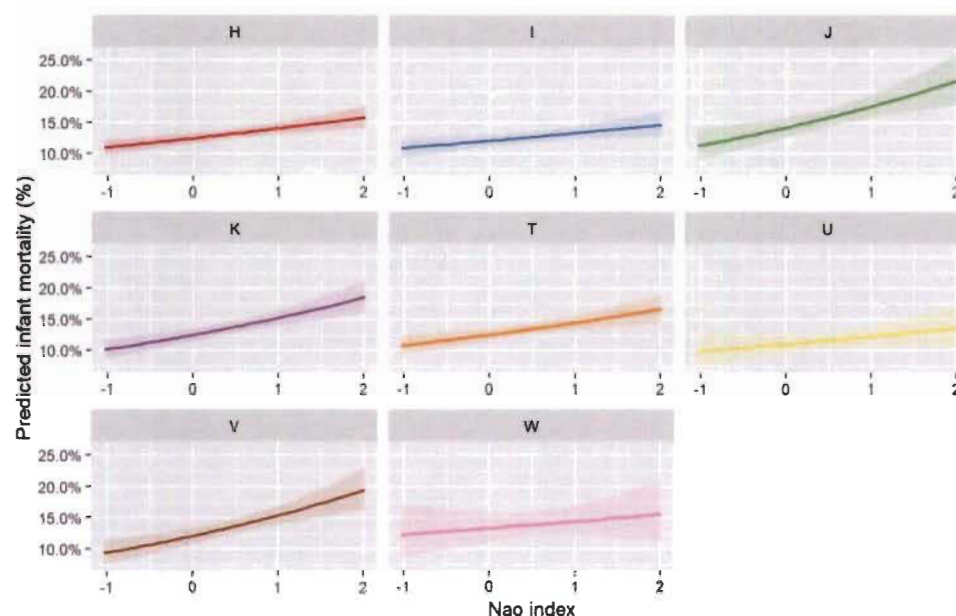


Figure S1 IM of mitochondrial haplogroup under the NAO gradients in the Saguenay–Lac-Saint-Jean population.

Each separate interaction from figure 13 for the NAO with their 95% confidence intervals.

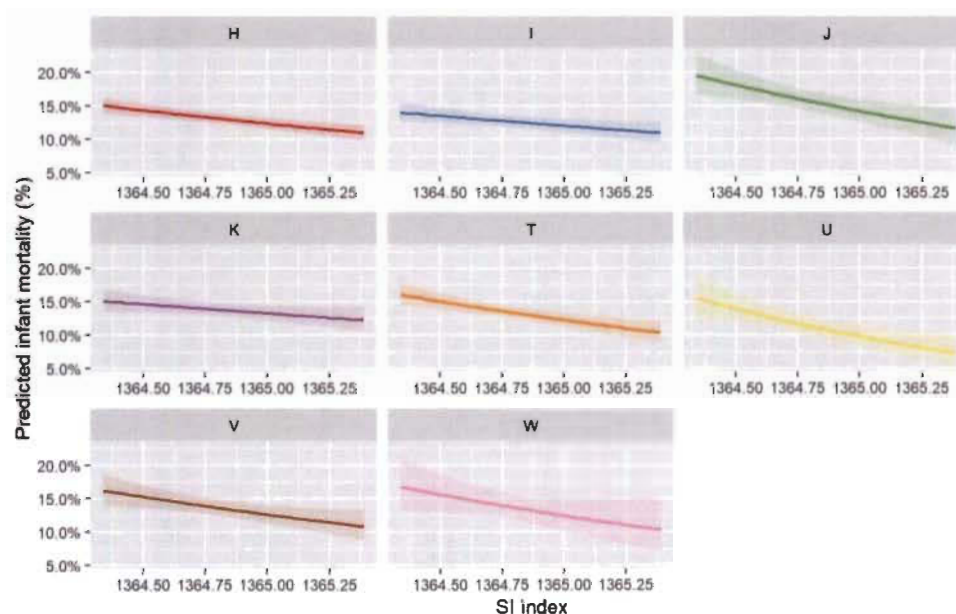


Figure S2 IM of mitochondrial haplogroup under the SI gradients in the Saguenay–Lac-Saint-Jean population.

Each separate interaction from figure 13 for the SI with their 95% confidence intervals.

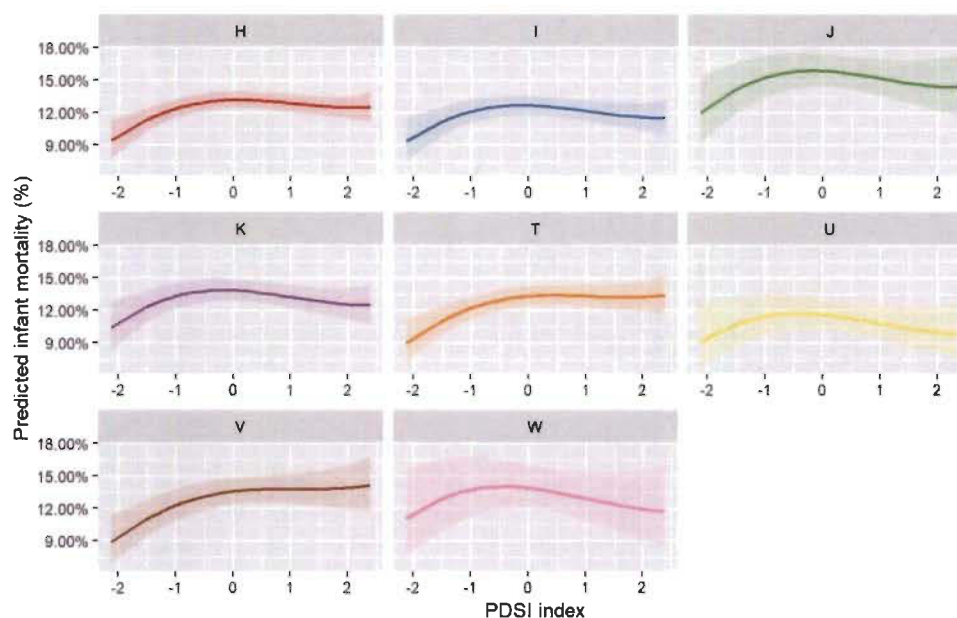


Figure S3 IM of mitochondrial haplogroup under the PDSI gradients in the Saguenay-Lac-Saint-Jean population.

Each separate interaction from figure 13 for the PDSI with their 95% confidence intervals.

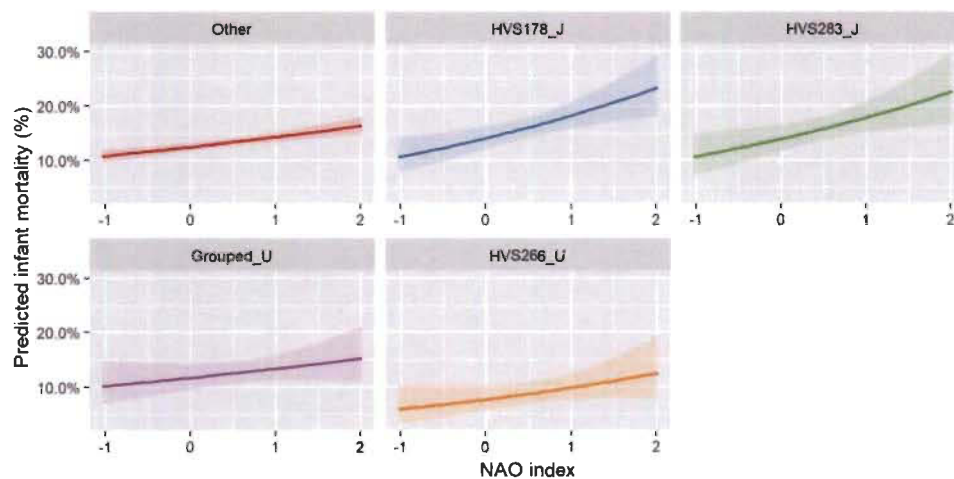


Figure S4 IM of mitochondrial haplotype under the NAO gradients in the Saguenay-Lac-Saint-Jean population.

Each separate interaction from figure 18 for the NAO with their 95% confidence intervals.

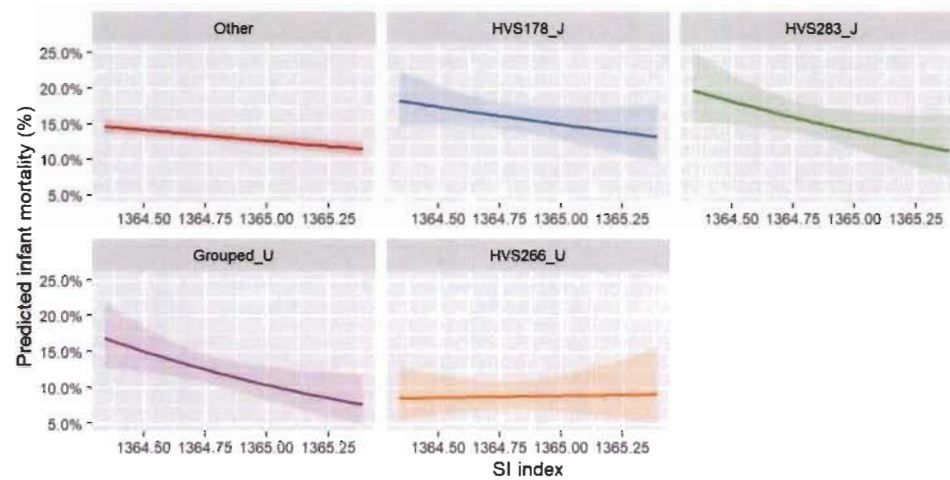


Figure S5 IM of mitochondrial haplotype under the SI gradients in the Saguenay--Lac-Saint-Jean population.
Each separate interaction from figure 18 for the SI with their 95% confidence intervals.

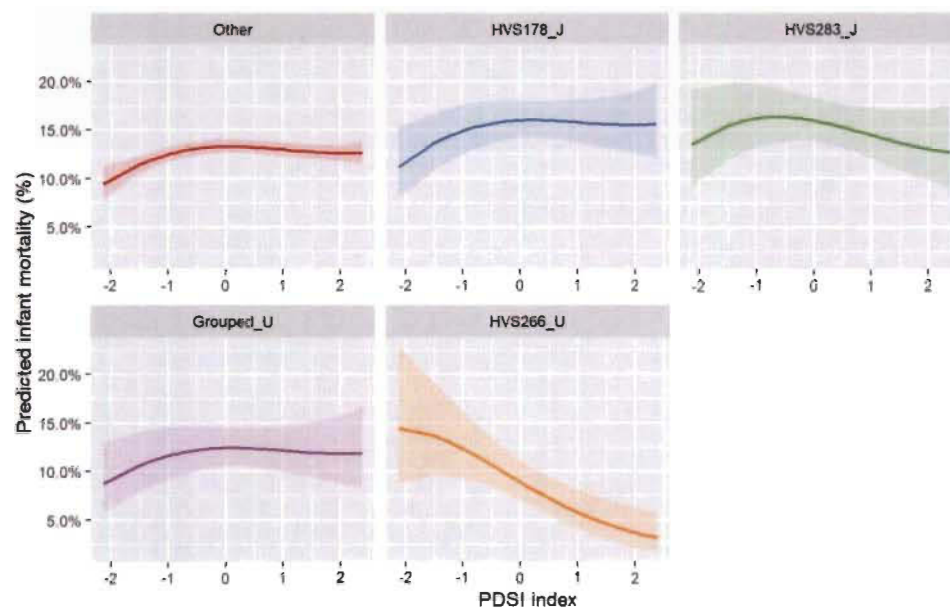


Figure S6 IM of mitochondrial haplotype under the PDSI gradients in the Saguenay-Lac-Saint-Jean population.
Each separate interaction from figure 18 for the PDSI with their 95% confidence intervals.

References

- Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, 23(2), 147-147. doi:10.1038/13779
- Ballard, J. W., Melvin, R. G., Katewa, S. D., & Maas, K. (2007). Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution*, 61(7), 1735-1747. doi:10.1111/j.1558-5646.2007.00133.x
- Ballard, J. W., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Mol Ecol*, 13(4), 729-744.
- Ballard, J. W., & Youngson, N. A. (2015). Review: can diet influence the selective advantage of mitochondrial DNA haplotypes? *Biosci Rep*, 35(6). doi:10.1042/BSR20150232
- Balloux, F., Handley, L. J., Jombart, T., Liu, H., & Manica, A. (2009). Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc Biol Sci*, 276(1672), 3447-3455. doi:10.1098/rspb.2009.0752
- Bergeron, P., Milot, E., Mayer, F. M., Boisvert, M., Réale, D., & Pelletier, F. (2014). Solar Irradiance, Survival and Longevity in a Pre-industrial Human Population. *Human Ecology*, 42(4), 645-650. doi:10.1007/s10745-014-9671-7
- Brommer, J. E., Rattiste, K., & Wilson, A. J. (2008). Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull *Larus canus*. *Proc Biol Sci*, 275(1635), 687-693. doi:10.1098/rspb.2007.0951
- Brown, C. R., Brown, M. B., Roche, E. A., O'Brien, V. A., & Page, C. E. (2016). Fluctuating survival selection explains variation in avian group size. *Proceedings of the National Academy of Sciences*, 113(18), 5113-5118. doi:10.1073/pnas.1600218113
- Bruckner, T. A., Gailey, S., Hallman, S., Amorevieta-Gentil, M., Dillon, L., & Gagnon, A. (2018). Epidemic cycles and environmental pressure in colonial Quebec. *American Journal of Human Biology*, 30(5), e23155. doi:10.1002/ajhb.23155
- Camus, M. F., Sgrò, C. M., Wolff, J. N., & Dowling, D. K. (2017). Experimental Support That Natural Selection Has Shaped the Latitudinal Distribution of Mitochondrial Haplotypes in Australian *Drosophila melanogaster*. *Mol Biol Evol*, 34(10), 2600-2612. doi:10.1093/molbev/msx184

- Camus, M. F., Wolf, J. B., Morrow, E. H., & Dowling, D. K. (2015). Single Nucleotides in the mtDNA Sequence Modify Mitochondrial Molecular Function and Are Associated with Sex-Specific Effects on Fertility and Aging. *Curr Biol*, 25(20), 2717-2722. doi:10.1016/j.cub.2015.09.012
- Chalmers, R. M., & Schapira, A. H. V. (1999). Clinical, biochemical and molecular genetic features of Leber's hereditary optic neuropathy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1410(2), 147-158. doi:https://doi.org/10.1016/S0005-2728(98)00163-7
- Charbonneau, H., Desjardins, B., Guillemette, A., Landry, Y., Legaré, J., & Nault, F. (1993). The first French Canadians: Pioneers in the St. Lawrence Valley. *Newark: University of Delaware Press*.
- Chen, A., Raule, N., Chomyn, A., & Attardi, G. (2012). Decreased reactive oxygen species production in cells with mitochondrial haplogroups associated with longevity. *PLoS One*, 7(10), e46473. doi:10.1371/journal.pone.0046473
- Christie, J. S., Castro, J. A., Oliver, P., Picornell, A., Ramon, M. M., & Moya, A. (2004). Fitness and life-history traits of the two major mitochondrial DNA haplotypes of *Drosophila subobscura*. *Heredity (Edinb)*, 93(4), 371-378. doi:10.1038/sj.hdy.6800513
- Currie, R. G., Wyatt, T., & O'Brien, D. P. (1993). Deterministic signals in european fish catches, wine harvests, and sea-level, and further experiments. *International Journal of Climatology*, 13(6), 665-687.
- Dato, S., Passarino, G., Rose, G., Altomare, K., Bellizzi, D., Mari, V., De Benedictis, G. (2004). Association of the mitochondrial DNA haplogroup J with longevity is population specific. *Eur J Hum Genet*, 12(12), 1080-1082. doi:10.1038/sj.ejhg.5201278
- Derosas, R. (2009). The joint effect of maternal malnutrition and cold weather on neonatal mortality in nineteenth-century Venice: an assessment of the hypothermia hypothesis. *Popul Stud (Camb)*, 63(3), 233-251. doi:10.1080/00324720903165449
- Doyon, A., Labuda, D., Moreau, C., Larmuseau, M., Vézina, H., Noël, Y., Milot, E. (submitted). Population-wide assessment of spatiotemporal variation in DNA lineage marker frequencies and implications for their use in forensic science. *Manuscript Draft*.
- Dunger, D. B., Petry, C. J., & Ong, K. K. (2006). Genetic Variations and Normal Fetal Growth. *Hormone Research in Paediatrics*, 65(suppl 3)(Suppl. 3), 34-40. doi:10.1159/000091504

- Feder, J., Ovadia, O., Blech, I., Cohen, J., Wainstein, J., Harman-Boehm, I., Mishmar, D. (2009). Parental diabetes status reveals association of mitochondrial DNA haplogroup J1 with type 2 diabetes. *BMC Medical Genetics*, 10(1), 60. doi:10.1186/1471-2350-10-60
- Feng, G. F., Zhang, J., Feng, L. M., Shen, N. X., Li, L. J., & Zhu, Y. M. (2013). Mitochondrial DNA haplogroup associated with sperm motility in the Han population. *Asian J Androl*, 15(5), 630-633. doi:10.1038/aja.2013.83
- Gagnon, A., & Mazan, R. (2009). Does exposure to infectious diseases in infancy affect old-age mortality? Evidence from a pre-industrial population. *Social Science & Medicine*, 68(9), 1609-1616. doi:<https://doi.org/10.1016/j.socscimed.2009.02.008>
- Garvin, M. R., Bielawski, J. P., Sazanov, L. A., & Gharrett, A. J. (2015). Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *Journal of Zoological Systematics and Evolutionary Research*, 53(1), 1-17. doi:10.1111/jzs.12079
- Gauvreau, D., & Bourque, M. (1988). Mouvements migratoires et familles : le peuplement du Saguenay avant 1911. *Revue d'histoire de l'Amérique française*, 42(2). doi:10.7202/304677ar
- Gershoni, M., Templeton, A. R., & Mishmar, D. (2009). Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays*, 31(6), 642-650. doi:10.1002/bies.200800139
- Grosbois, V., & Thompson, P. M. (2005). North Atlantic Climate Variation Influences Survival in Adult Fulmars. *Oikos*, 109(2), 273-290.
- Hamel, M. (1993). De Charlevoix au Saguenay : caractéristiques des familles émigrantes au XIX siècle. *Revue d'histoire de l'Amérique française*, 47(1). doi:10.7202/305180ar
- Hurrell, J. W. (1995). Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Science*, 269(5224), 676-679. doi:10.1126/science.269.5224.676
- Kodera, K. (2002). Solar cycle modulation of the North Atlantic Oscillation: Implication in the spatial structure of the NAO. *Geophysical Research Letters – GEOPHYS RES LETT*, 29. doi:10.1029/2001GL014557
- Landry, Y., & Lessard, R. (2008). Les causes de décès aux XVIIe et XVIIIe siècles d'après les registres paroissiaux québécois. *Revue d'histoire de l'Amérique française*, 48(4), 509-526. doi:10.7202/305363ar

- Lean, J. (2000). Evolution of the Sun's Spectral Irradiance Since the Maunder Minimum. *Geophysical Research Letters*, 27(16), 2425-2428. doi:10.1029/2000gl000043
- Leng, G., & Hall, J. (2019). Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. *Science of The Total Environment*, 654, 811-821. doi:https://doi.org/10.1016/j.scitotenv.2018.10.434
- Lummaa, V. (2003). Early developmental conditions and reproductive success in humans: downstream effects of prenatal famine, birthweight, and timing of birth. *Am J Hum Biol*, 15(3), 370-379. doi:10.1002/ajhb.10155
- Malik, S., Diot, A., Morten, K., Dombi, E., Vatish, M., Boyd, C. A. R., & Poulton, J. (2017). Acute nutritional stress during pregnancy affects placental efficiency, fetal growth and adult glucose homeostasis. *Oncotarget*, 8(65), 109478-109486. doi:10.18632/oncotarget.22695
- McKenney, W. M. (2001). Understanding the neonatal immune system: high risk for infection. *Crit Care Nurse*, 21(6), 35-47.
- Milot, E., Moreau, C., Gagnon, A., Cohen, A. A., Brais, B., & Labuda, D. (2017). Mother's curse neutralizes natural selection against a human genetic disease over three centuries. *Nat Ecol Evol*, 1(9), 1400-1406. doi:10.1038/s41559-017-0276-6
- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Wallace, D. C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A*, 100(1), 171-176. doi:10.1073/pnas.0136972100
- Moore, V. M., Davies, M. J., Willson, K. J., Worsley, A., & Robinson, J. S. (2004). Dietary composition of pregnant women is related to size of the baby at birth. *J Nutr*, 134(7), 1820-1826. doi:10.1093/jn/134.7.1820
- Moreau, C., Bherer, C., Vezina, H., Jomphe, M., Labuda, D., & Excoffier, L. (2011). Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science*, 334(6059), 1148-1150. doi:10.1126/science.1212880
- Moreau, C., Vézina, H., Jomphe, M., Lavoie, È.-M., Roy-Gagnon, M.-H., & Labuda, D. (2011). When Genetics and Genealogies Tell Different Stories—Maternal Lineages in Gaspesia. *Ann Hum Genet*, 75(2), 247-254. doi:10.1111/j.1469-1809.2010.00617.x
- Moreau, C., Vezina, H., Yotova, V., Hamon, R., de Knijff, P., Sinnett, D., & Labuda, D. (2009). Genetic heterogeneity in regional populations of Quebec-parental lineages in the Gaspé Peninsula. *Am J Phys Anthropol*, 139(4), 512-522. doi:10.1002/ajpa.21012

- Nielsen, R., Hubisz, M. J., Hellmann, I., Torgerson, D., Andrés, A. M., Albrechtsen, A., Clark, A. G. (2009). Darwinian and demographic forces affecting human protein coding genes. *Genome research*, 19(5), 838-849. doi:10.1101/gr.088336.108
- Ortega, P., Lehner, F., Swingedouw, D., Masson-Delmotte, V., Raible, C. C., Casado, M., & Yiou, P. (2015). A model-tested North Atlantic Oscillation reconstruction for the past millennium. *Nature*, 523(7558), 71-74. doi:10.1038/nature14518
- Ottersen, G., Planque, B., Belgrano, A., Post, E., Reid, P. C., & Stenseth, N. C. (2001). Ecological effects of the North Atlantic Oscillation. *Oecologia*, 128(1), 1-14. doi:10.1007/s004420100655
- Polgreen, P. M., & Polgreen, E. L. (2018). Infectious Diseases, Weather, and Climate. *Clin Infect Dis*, 66(6), 815-817. doi:10.1093/cid/cix1105
- Pustilnik, L. A., & Yom Din, G. (2004). Influence of solar activity on state of wheat market in medieval England. *Solar Physics*, 223(1-2), 335-356.
- Raby, B. A., Klanderman, B., Murphy, A., Mazza, S., Camargo, C. A., Jr., Silverman, E. K., & Weiss, S. T. (2007). A common mitochondrial haplogroup is associated with elevated total serum IgE levels. *J Allergy Clin Immunol*, 120(2), 351-358. doi:10.1016/j.jaci.2007.05.029
- Ramírez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F., & Navarro, A. (2008). Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*, 179(1), 555-567. doi:10.1534/genetics.107.083006
- Robert, A., Toupance, B., Tremblay, M., & Heyer, E. (2009). Impact of inbreeding on fertility in a pre-industrial population. *Eur J Hum Genet*, 17(5), 673-681. doi:10.1038/ejhg.2008.237
- Rocklov, J., Edvinsson, S., Arnqvist, P., de Luna, S. S., & Schumann, B. (2014). Association of seasonal climate variability and age-specific mortality in northern Sweden before the onset of industrialization. *Int J Environ Res Public Health*, 11(7), 6940-6954. doi:10.3390/ijerph110706940
- Ross, J. M., Coppotelli, G., Hoffer, B. J., & Olson, L. (2014). Maternally transmitted mitochondrial DNA mutations can reduce lifespan. *Sci Rep*, 4, 6569. doi:10.1038/srep06569
- Rousset, F., & Ferdy, J.-B. (2014). Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, 37(8), 781-790. doi:10.1111/ecog.00566

- Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., & Wallace, D. C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, 303(5655), 223-226. doi:10.1126/science.1088434
- Schumann, B., Edvinsson, S., Evengard, B., & Rocklov, J. (2013). The influence of seasonal climate variability on mortality in pre-industrial Sweden. *Glob Health Action*, 6, 20153. doi:10.3402/gha.v6i0.20153
- Scott, S., Duncan, S. R., & Duncan, C. J. (1995). Infant mortality and famine: a study in historical epidemiology in northern England. *Journal of epidemiology and community health*, 49(3), 245-252. doi:10.1136/jech.49.3.245
- Séguin, N. (1976). L'économie agro-forestière : genèse du développement au Saguenay au XIX siècle. *Revue d'histoire de l'Amérique française*, 29(4). doi:10.7202/303486ar
- Sigurgardottir, S., Helgason, A., Gulcher, J. R., Stefansson, K., & Donnelly, P. (2000). The mutation rate in the human mtDNA control region. *Am J Hum Genet*, 66(5), 1599-1609.
- Skjærvø, G. R., Fossøy, F., & Røskoft, E. (2015). Solar activity at birth predicted infant survival and women's fertility in historical Norway. *Proc Biol Sci*, 282(1801), 20142032-20142032. doi:10.1098/rspb.2014.2032
- Slatkin, M., & Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, 129(2), 555-562.
- Sorg, M., & Craig, B. (1983). Patterns of Infant Mortality in the Upper St. John Valley French Population: 1791-1838. *Human biology*, 55, 100-113.
- Torroni, A., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., Scozzari, R. (2001). Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet*, 69(6), 1348-1356. doi:10.1086/324511
- Wallace, D. C. (2015). Mitochondrial DNA variation in human radiation and disease. *Cell*, 163(1), 33-38. doi:10.1016/j.cell.2015.08.067
- Wettstein, J. J., & Mearns, L. O. (2002). The Influence of the North Atlantic-Arctic Oscillation on Mean, Variance, and Extremes of Temperature in the Northeastern United States and Canada. *Journal of Climate*, 15(24), 3586-3600. doi:10.1175/1520-0442(2002)015<3586:TIOTNA>2.0.CO;2

Zhang, Z., Mann, M. E., & Cook, E. R. (2004). Alternative methods of proxy-based climate field reconstruction: application to summer drought over the conterminous United States back to AD 1700 from tree-ring data. *The Holocene*, 14(4), 502-516. doi:10.1191/0959683604hl727rp

CHAPITRE III

DISCUSSION ET PERSPECTIVE

La méthode d'imputation de marqueurs haploïdes nous a permis d'imputer à une population historique un nombre impressionnant d'haplotypes mitochondriaux avec un bon degré de précision. Cela a permis de tester l'hypothèse que le climat a un impact sur la répartition des fréquences de l'ADNmt chez l'humain en tant qu'agent de sélection sur le génome mitochondrial. Ce test fut fait en mesurant, par le biais d'indicateurs climatiques à grande échelle, l'effet du climat sur la mortalité infantile associée à chacun des mitotypes dans une population humaine. Dans un premier temps, j'ai démontré que la variabilité dans le taux de mortalité infantile de la population du Saguenay-Lac-Saint-Jean s'explique en bonne partie par les variables climatiques sélectionnées et donc, que ces variables contenant un signal de l'effet présumé du climat sur chacun des mitotypes. En un deuxième temps, j'ai mis au jour une relation entre le mitotype et la mortalité infantile dans cette même population, d'abord sans considération du climat. Ceci confirme que la mortalité infantile 0-1 an était un bon choix comme composante de fitness pour ensuite étudier l'interaction entre le climat et le mitotype. Deux mitotypes sont ressortis comme étant sous sélection à l'époque préindustrielle. L'un sous sélection positive (U) et l'autre sous sélection négative (J). L'ajoute de l'interaction mitotype \times climat a révélé une relation positive entre les haplogroupes I, V et K et l'Oscillation nord-atlantique, ainsi qu'une relation négative entre les haplogroupes J et U et l'irradiance solaire.

Nous avons obtenu des résultats appuyant les hypothèses énoncées au chapitre I, à savoir que le climat peut effectivement créer les conditions propices pour engendrer de la sélection au niveau de l'ADNmt en modulant la variance dans le taux de mortalité (plus grande variance dans la mortalité infantile dans les extrêmes climatiques défavorables que dans les extrêmes favorables).

Les effets du NAO et du SI et leur interaction appuient l'hypothèse à l'effet que la rigueur du climat et la disponibilité des ressources pouvant en résulter affectent le fitness des individus. Les variables climatiques utilisées ne permettent pas de démontrer une relation causale, mais le NAO est connu pour être un bon prédicteur de la rigueur de l'hiver dans plusieurs cas et le SI, de la disponibilité des ressources. Considérant les fenêtres temporelles retenues pour le NAO, soit la moyenne de l'année en cours et de l'année précédente, on peut penser que si pendant deux années consécutives l'hiver a été long et ardu, ce sont deux années où le temps des récoltes a pu être repoussé (printemps tardif). Pendant ce temps, les individus ont été soumis à des conditions climatiques plus difficiles et possiblement nécessité plus de ressources alimentaires. Si ce phénomène se combine à une année de sécheresse menant à de récoltes moins abondantes, cela peut rendre les enfants davantage exposés aux maladies et augmenter leur mortalité. De plus, l'effet du SI et du NAO était amplifié ou réduit selon le type d'ADNmt, ce qui concorde avec les effets rapportés sur l'apport faible ou déséquilibré en nutriments comme le ratio protéines : carbohydrate qui a une influence sur le fitness d'espèce animale modulé par leur génotype mito (Aw et al., 2017; Sun et al., 2014; Towarnicki et al., 2017) ou les effets de stress qui ont un impact direct au niveau cellulaire (Baker et al., 2019).

Quelques bémols sont toutefois à mentionner au niveau de ces analyses. Dans l'évaluation de la mortalité infantile, j'ai présumé que les personnes n'ayant pas de date de décès avaient survécu après l'âge de 1 an. Cependant, il est connu que dans le Québec préindustriel, il y avait un certain sous-enregistrement des décès infantiles qui variait dans le temps (Gagnon et al., 2009). Cela a pu avoir comme effet de réduire l'estimation du vrai taux de mortalité infantile. De plus, lors de la modélisation, je n'ai pas pris en compte certains paramètres pouvant causer de la variabilité au niveau de la mortalité. C'est entre autres le cas de ceux identifiés à la section 1.10, comme les intervalles protogénésique, intergénésique et l'âge de la mère au moment de la naissance d'un enfant. Pour les intervalles, la raison était pour limiter davantage l'attrition du jeu de données, afin de ne pas réduire la couverture mitochondriale pour avoir le maximum d'haplogroupes différents avec un nombre d'individus suffisant grand pour chacun. En ce qui concerne l'âge de la mère, le problème en était un de colinéarité. En effet,

cette variable est fortement corrélée avec le rang de naissance. Dans ce type de modèle, lorsque des variables explicatives sont trop corrélées, l'attribution de variance ne se fait pas bien et la stabilité du modèle en souffre. Il y a aussi probablement une corrélation entre le rang de naissance et l'intervalle intergénéral.

Nous avons donc gardé le rang de naissance, car les données étaient plus complètes. Avoir mis de côté ces variables a peut-être quelque peu causé une surévaluation des effets mitochondriaux estimés. Toutefois, le fait d'omettre ces variables ne devrait pas avoir biaisé les résultats dans un sens en particulier. À moins que certains variants mitochondriaux aient un réel impact sur la fécondité. Dans ce cas, les variants augmentant la fécondité auraient des intervalles plus courts (en moyenne) et un taux de mortalité un peu plus élevé alors que ceux diminuant la fécondité aurait des intervalles plus longs, avec un risque moins grand de mortalité infantile. Ce qui demeurerait alors un effet du mitotype.

Finalement, je n'ai pas évalué l'effet du background nucléaire de même que les interactions mito-nucléaires, pouvant là aussi laisser plus de place aux effets mitochondriaux seuls, ou au contraire introduire du bruit limitant la capacité à détecter le signal de sélection.

Ce projet incluait plusieurs défis. L'un d'entre eux a été de trouver quel indicateur de fitness serait le mieux adapté afin d'évaluer la sélection sur l'ADNmt en lien avec le climat dans une population humaine. Comme il a été mentionné à la section 1.9, ce choix s'est arrêté sur la mortalité infantile, à la suite d'analyses préliminaires sur plusieurs autres mesures, non rapportées ici. Il y a beaucoup de référents dans la littérature à propos de l'effet du climat sur la fécondité ainsi que sur les effets possibles de la variation mitochondriale sur ce trait d'histoire de vie (Camus et al., 2015; Otten et al., 2015; Skjærvø et al., 2015). Toutefois, lorsqu'est venu le temps de modéliser ce trait, de nombreux défis se sont présentés à moi. La distribution statistique de la fécondité dans la population du Saguenay-Lac-Saint-Jean présente une bimodalité avec un excès de 0 (Figure 3.1).

J'ai effectué de nombreux tests pour trouver la meilleure façon de modéliser cette distribution (*zero-inflated* Poisson, *zero-inflated* binomiale, modèle Hurdle, etc.) afin de ne pas prendre seulement les individus avec une fécondité > 0 , ce qui aurait mené à une attrition importante des données et de la couverte mitochondriale. De plus, il s'est avéré compliqué de déterminer la fenêtre temporelle pertinente pour caractériser le climat auquel un individu a été exposé avant et pendant sa vie féconde. De plus, nous ne pouvions pas prendre en compte ceux qui n'avaient pas eu d'enfant pour faire l'évaluation, car il n'y a aucune variable climatique pouvant s'associer au fait de ne pas avoir d'enfant.

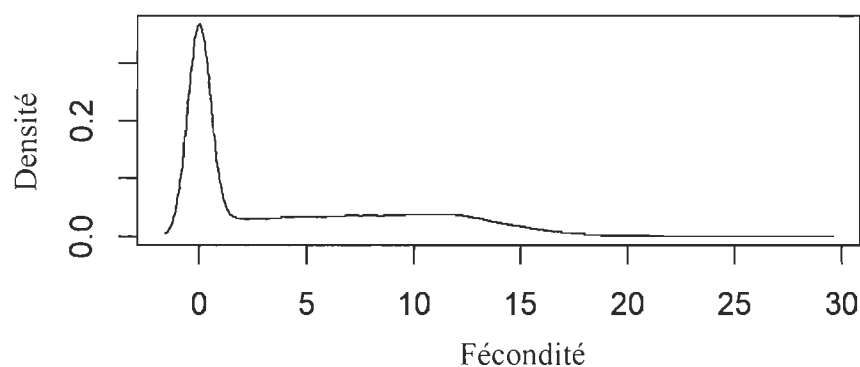


Figure 3.1 Distribution de la fécondité dans la population du Saguenay-Lac-Saint-Jean.

Pour pallier ces problèmes liés à la mesure de fécondité, j'ai utilisé une mesure de fitness, dite « démographique » selon la méthode décrite dans Moorad (2014). Cette mesure, beaucoup plus complète, intègre la survie juvénile, l'appariement (mariage) et la fécondité d'un individu, et ce, de façon relative à la croissance de la population, qui peut varier entre les cohortes. Le fitness est calculé par cohorte et chaque individu se voit attribuer une seule mesure de fitness relatif, c'est-à-dire par rapport à la moyenne de sa cohorte. J'ai donc ajusté des modèles en utilisant cette mesure de fitness démographique, en lien avec le climat et l'ADNmt. Plusieurs études montrent d'ailleurs l'impact de l'environnement de naissance sur le développement et comment celui-ci peut influencer la vie adulte et la reproduction (Burton et al., 2014). Toutefois, cette mesure de fitness avait aussi une distribution présentant les problèmes observés avec la fécondité, d'une part à cause du haut taux de mortalité avant l'âge à la reproduction et d'autre part, à cause de la perte d'information sur les individus ayant quitté la région du SLSJ avant de

se reproduire. En raison des nombreux défis d'analyse, il faudrait explorer ces questions dans un projet séparé. Les estimations avec le fitness démographique permettraient de prédire plus directement le changement de fréquence des mitotypes d'une génération à l'autre dû à la sélection.

J'ai aussi tenté de modéliser l'effet du climat sur la survie juvénile (1-15 ans) à l'aide du modèle à risque proportionnel de Cox. Cette approche semblait très prometteuse, car chaque pas de temps est traité comme indépendant et permet d'évaluer le risque de mortalité à chaque âge en lien avec le climat et l'ADNmt. Aussi, avec cette approche, j'aurais pu limiter l'attrition du jeu de données comme c'était le cas des analyses sur la fécondité et le fitness démographique, car on peut prendre en compte les données tronquées (*right censoring*), telles que celles pour les individus qui quittaient le territoire à l'âge adulte. J'ai obtenu des résultats préliminaires avec ces analyses. Cependant, ce type de modèle ne permet pas d'intégrer tous les paramètres que nous voulions prendre en compte pour mieux isoler la variabilité dans le fitness due à la variation mitochondriale (p. ex. l'autocorrélation spatiale, l'environnement maternel et la variance génétique additive). C'est pourquoi, j'ai finalement arrêté mon choix sur la mortalité infantile comme composante de fitness pour mesurer la sélection naturelle. Cette variable limitait la perte de données et permettait une bonne évaluation des conditions climatiques avec les séries temporelles que j'ai utilisées.

Une autre grande partie de mon travail a consisté à adapter la méthode d'imputation probabiliste développée par Doyon et al. (soumis) dans le cadre de sa maîtrise sur la dynamique des marqueurs génétiques liés au sexe. Cette méthode a été développée à la base pour attribuer des marqueurs du chromosome Y aux individus de la base de données BALSAC. J'ai donc adapté la méthode (et les scripts R) pour l'appliquer aux haplotypes mitochondriaux et les attribuer aux individus de cette même base de données. Ces résultats ont été ajoutés à l'article d'Alexandra Doyon, en processus de soumission, sur lequel je figure comme co-auteur.

La grande différence entre l'ADNmt et le chromosome Y est que dans le cas des marqueurs Y, lorsqu'il y avait une mutation, on considérait automatiquement qu'elle donnait un nouvel haplotype. Donc, lorsqu'une mutation était simulée dans le pedigree, l'individu sur lequel tombait cette mutation ainsi que l'ensemble de ses descendant mâles se voyaient attribuer un haplotype désigné comme étant inconnu. Dans notre cas, lorsqu'il n'y avait qu'un seul haplotype de séquencé par lignée, nous appliquions cette règle. Cependant, il arrivait souvent que nous ayons 2, 3 ou même 4 haplotypes séquencés par lignée, quelquefois avec seulement 1 ou 2 nucléotides de différence dans la région de contrôle. Nous devons donc prendre en compte la possibilité que cette différence soit issue d'une erreur généalogique, mais aussi de mutations de l'haplotype de la fondatrice de la lignée, qui aurait pu muter successivement plusieurs fois compte tenu du taux de mutation élevé de la région de contrôle. Cette façon de faire a permis de savoir, avec une bonne approximation, à quel endroit dans la généalogie l'haplotype mitochondrial avait muté et de détecter de probables erreurs généalogiques avec une bonne précision. Je n'ai toutefois pas intégré la possibilité qu'il y ait des « back mutations », facilitant peut-être le match entre deux haplotypes se différenciant de quelques SNPs et pouvant orienter l'assignation dans certaines branches de la généalogie. L'assignation s'est faite en tenant compte que les mutations s'accumulaient sur l'ADNmt. Donc, si deux haplotypes avaient les mutations suivantes : H1 (A73G et C150T), H2 (A73G, C150T, A263G et A750G), l'haplotype H1 avait la possibilité de muter deux fois pour devenir H2 mais H2 ne pouvait devenir H1.

En complément des haplotypes mitochondriaux à disposition, nous avons fait séquencer le génome complet de 95 des échantillons. La sélection des haplotypes à séquencer s'est faite selon la représentativité des lignées maternelles dans la base de données BALSAC. Un jumelage préalable des haplotypes défini par les SNPs présents sur les régions de contrôle avait été fait par Claudia Moreau du CHU Sainte-Justine. Nous avons donc calculé quelles lignées maternelles associées à ces haplotypes offraient la plus grande couverture possible (nombre d'individus), à la fois sur l'ensemble du Québec et au Saguenay-Lac-Saint-Jean. Le séquençage a été effectué au Centre de génomique de l'Université de Pompeu Fabra à Barcelone par notre collaborateur

Ferran Casals. La stratégie utilisée est la réplication du génome à partir de deux grands amplicons par PCR, en utilisant une polymérase spécifique pour les fragments longs (LA Taq de Takara Bio Industries). Cette polymérase combine l'ADN polymérase Taq et une polymérase de correction de l'ADN. Les banques de gènes ont été construites avec Nextera XT et la qualité vérifiée par Bionalyzer. Une normalisation des concentrations a été effectuée après une quantification avec la méthode du fluorophore (PicoGreen) intercalé et le séquençage s'est effectué par MiSeq (2 x 250). L'entièreté des SNPs retrouvés dans les 95 séquences a été analysée avec une approche d'étude d'association pangénomique (GWAS) afin de trouver de possibles mutations causales pouvant expliquer les effets observés des haplogroupes/haplotypes mitochondriaux sur la mortalité infantile. Pour cette analyse, seuls les polymorphismes avec un score de qualité de Phred supérieur à 30 ont été pris en compte; équivalent à la probabilité d'un appel de base incorrect d'une précision de 1 sur 1 000. (Ewing et al., 1998). Les positions 300 à 315 ont été exclues, car elles correspondent à une région homopolymère de cytosine, connue pour être problématique à séquencer, ce qui a été confirmé par nos données de séquençage brutes.

Le faible nombre de SNPs ayant une représentation suffisante (portés par au moins 100 personnes) n'a pas permis une résolution suffisante pour trouver une région causale candidate sur le génome. La figure 3.2 montre les résultats obtenus pour l'analyse GWAS sous forme de Manhattan plot circulaire. Le cercle multicolore au centre représente le génome mitochondrial. Chacun des gènes y sont représentés avec une couleur différente, ainsi que l'ARN ribosomique en rouge, l'ARN de transfert en noir et les régions non codantes en orange. Les bases incrémentent en sens horaire en commençant par la position « midi ». Chacun des points observés représente l'association d'un SNPs avec la mortalité infantile ($-\log_{10}$ p-value). Plus un point s'éloigne du centre, plus l'association est forte. On voit bien sur cette figure, que la région la plus dense est la région de contrôle, là où la variabilité est la plus grande. De plus, certains d'entre eux ne sont pas indépendants (présents sur une séquence seulement) mais transmis en bloc (accompagnés par d'autres SNP aussi présent seulement sur cette séquence) et nous n'avons pas été en mesure d'isoler leur valeur d'association avec la mortalité infantile des autres SNPs liés.

Cependant, en séparant les mutations silencieuses des mutations provoquant un changement d'acide aminé, nous avons quand même réussi à trouver des SNPs intéressants qui atteignent la valeur P de la correction de Bonferroni pour des niveaux significatifs suggestifs de 5 % (0.05 divisé par le nombre de SNPs, ici 327). Nous avons aussi reconstruit les séquences complètes à partir des fichiers de séquençage à l'aide du programme VCFtools et exécuté quelques tests avec DnaSP, un logiciel permettant l'analyse des sites polymorphes à partir d'un alignement de séquence. Nous avons notamment identifié les régions conservées sur le génome et réalisé des tests de diversité génétique en fenêtre coulissante comme le test de Tajima D (Tajima, 1989) et Fu & Li D* et F (Fu et al., 1993) dans le but de trouver des corrélations entre nos mesures de sélection et les régions sous sélection révélées par ces tests. Toutefois, ces résultats sont encore trop préliminaires pour en faire une description complète.

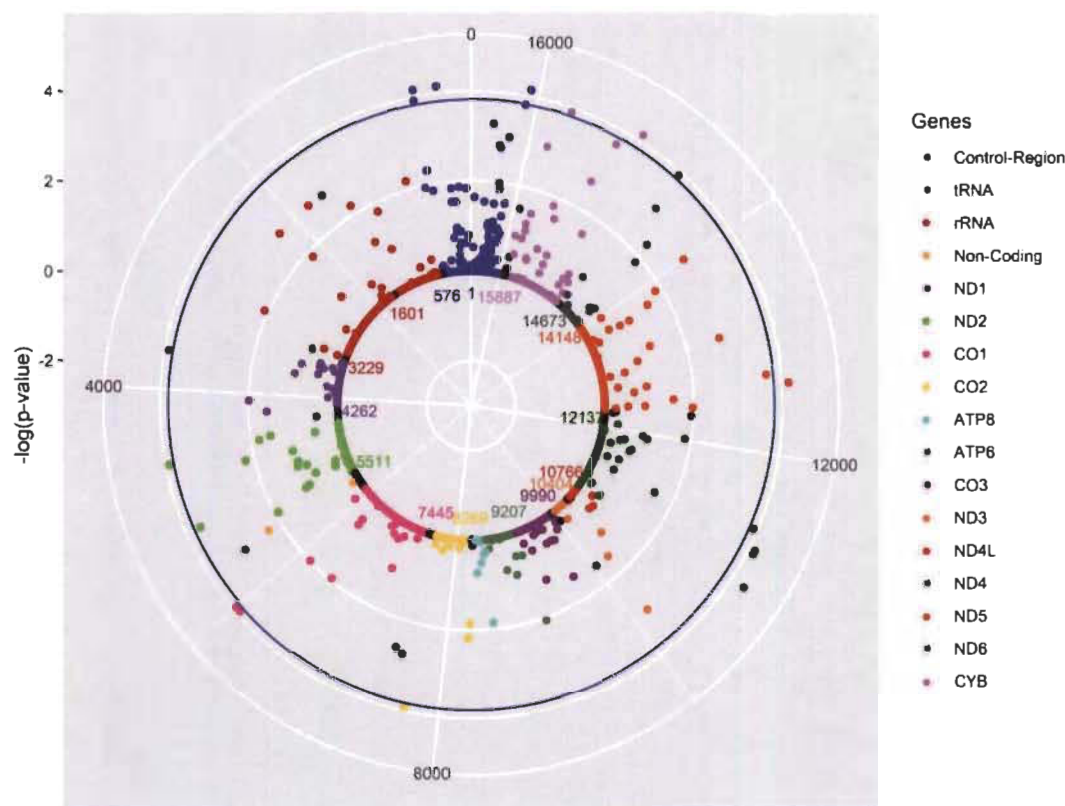


Figure 3.2 Manhattan plot circulaire de l'association des SNPs mitochondriaux avec la mortalité infantile au Saguenay-Lac-Saint-Jean.

La ligne bleue est le seuil corrigé de Bonferroni pour des niveaux significatifs au seuil $\alpha = 0.05$. Les codes R pour réaliser le graphique sont disponibles à <https://github.com/stephenturner/solarplot>

Perspective

Il serait intéressant, dans le futur, de refaire l'analyse sur ce même jeu de données, mais avec un filtre un peu plus détaillé, prenant en compte les variables laissées de côté ici, en plus de contrôler pour le background génétique nucléaire en intégrant la matrice d'apparentement génétique entre les individus, fournie par le pedigree. Une façon serait d'utiliser la matrice de covariance de parenté pour quantifier l'épistasie mito-nucléaire ou de développer un modèle utilisant le « gene dropping » où deux allèles sont attribués aux fondateurs et les génotypes des descendants sont générés. Ce qui serait aussi intéressant serait de faire le lien entre les individus typés de cette base de données et le Registre de population du Québec ancien (Dillon et al. 2018). Cette base de données contient des informations sur les naissances et décès de la fondation jusqu'en 1800. La période couverte étant beaucoup plus longue, il y aurait une étendue de valeurs des variables climatiques beaucoup plus intéressante que sur la période couverte au Saguenay-Lac-Saint-Jean. Avant de commencer à travailler avec les données du SLSJ, nous avons effectué quelques analyses préliminaires sur toute la période couverte avec BALSAC et avons observé des changements de fréquences beaucoup plus importants que ceux observés au SLSJ. De plus, il était possible de voir des trajectoires de fitness différentes selon les haplotypes mitochondriaux (Figure 3.3). Ce fitness avait été mesuré avec le nombre de descendants mariés d'un individu étant donné que seulement les mariages y sont recensés. Une autre question intéressante, serait de voir si le fitness change selon le sexe et tester l'hypothèse de la malédiction maternelle (*mother's curse*; Gemmell et al., 2004). Cette théorie prédit que les mutations qui s'accumulent sur l'ADNmt au fil des générations auront, en moyenne, un effet délétère plus important chez les mâles que chez les femelles.

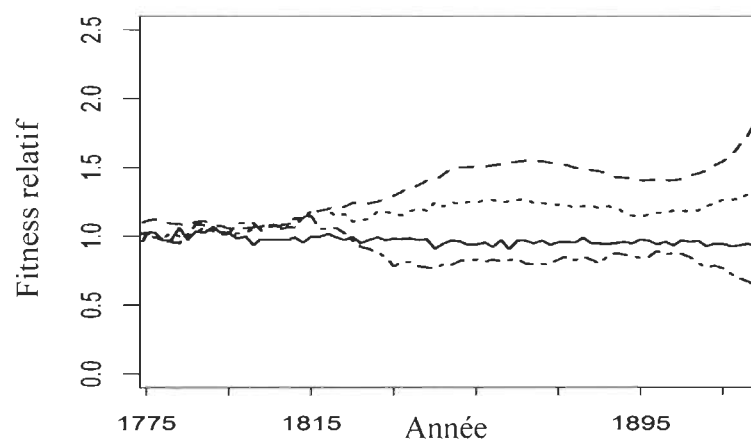


Figure 3.3 Différentes trajectoires de fitness temporelle pour 4 marqueurs d'ADNmt de la région de contrôle 1, de 1775 à 1915.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Amo, T., & Brand, M. D. (2007). Were inefficient mitochondrial haplogroups selected during migrations of modern humans? A test using modular kinetic analysis of coupling in mitochondria from cybrid cell lines. *The Biochemical journal*, 404(2), 345-351. doi:10.1042/BJ20061609
- Amo, T., Yadava, N., Oh, R., Nicholls, D. G., & Brand, M. D. (2008). Experimental assessment of bioenergetic differences caused by the common European mitochondrial DNA haplogroups H and T. *Gene*, 411(1-2), 69-76. doi:10.1016/j.gene.2008.01.007
- Amorevieta-Gentil, M. (2009). *Les niveaux et les facteurs déterminants de la mortalité infantile en Nouvelle-France et au début du Régime Anglais (1621-1779)* (Ph. D), Université de Montréal Les niveaux et les facteurs déterminants de la mortalité infantile en Nouvelle-France et au début du Régime Anglais (1621-1779) par Marilyn Amorevieta-Gentil Département de Démographie, Faculté des Arts et des Sciences.
- Aw, W. C., Garvin, M. R., Melvin, R. G., & Ballard, J. W. O. (2017). Sex-specific influences of mtDNA mitotype and diet on mitochondrial functions and physiological traits in *Drosophila melanogaster*. *PLoS One*, 12(11), e0187554-e0187554. doi:10.1371/journal.pone.0187554
- Baker, N., Patel, J., & Khacho, M. (2019). Linking mitochondrial dynamics, cristae remodeling and supercomplex formation: How mitochondrial structure can regulate bioenergetics. *Mitochondrion*, 49, 259-268. doi:10.1016/j.mito.2019.06.003
- Ballard, J. W., & Kreitman, M. (1995). Is mitochondrial DNA a strictly neutral marker? *Trends Ecol Evol*, 10(12), 485-488. doi:https://doi.org/10.1016/S0169-5347(00)89195-8
- Ballard, J. W., Melvin, R. G., Katewa, S. D., & Maas, K. (2007). Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution*, 61(7), 1735-1747. doi:10.1111/j.1558-5646.2007.00133.x
- Ballard, J. W., & Youngson, N. A. (2015). Review: can diet influence the selective advantage of mitochondrial DNA haplotypes? *Biosci Rep*, 35(6). doi:10.1042/BSR20150232

- Ballard, J. W. O., & Pichaud, N. (2014). Mitochondrial DNA: more than an evolutionary bystander. *Functional Ecology*, 28(1), 218-231. doi:10.1111/1365-2435.12177
- Balloux, F., Handley, L. J., Jombart, T., Liu, H., & Manica, A. (2009). Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc Biol Sci*, 276(1672), 3447-3455. doi:10.1098/rspb.2009.0752
- BALSAC. Fichier de population. Retrieved from <http://balsac.uqac.ca/>
- Bell, G. (2010). Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537), 87-97. doi:10.1098/rstb.2009.0150
- Bergamini, C. M., Gambetti, S., Dondi, A., & Cervellati, C. (2004). Oxygen, reactive oxygen species and tissue damage. *Curr Pharm Des*, 10(14), 1611-1626. doi:10.2174/1381612043384664
- Bergeron, P., Milot, E., Mayer, F. M., Boisvert, M., Réale, D., & Pelletier, F. (2014). Solar Irradiance, Survival and Longevity in a Pre-industrial Human Population. *Human Ecology*, 42(4), 645-650. doi:10.1007/s10745-014-9671-7
- Bourgeois-Pichat, J. (1951). La mesure de la mortalité infantile. Principes et méthodes. *Population*, 233-248.
- Brand, M. D., & Nicholls, D. G. (2011). Assessing mitochondrial dysfunction in cells. *The Biochemical journal*, 435(2), 297-312. doi:10.1042/BJ20110162
- Burton, T., & Metcalfe, N. B. (2014). Can environmental conditions experienced in early life influence future generations? *Proc Biol Sci*, 281(1785), 20140311-20140311. doi:10.1098/rspb.2014.0311
- Busiello, R. A., Savarese, S., & Lombardi, A. (2015). Mitochondrial uncoupling proteins and energy metabolism. *Frontiers in physiology*, 6, 36-36. doi:10.3389/fphys.2015.00036
- Camus, M. F., Sgrò, C. M., Wolff, J. N., & Dowling, D. K. (2017). Experimental Support That Natural Selection Has Shaped the Latitudinal Distribution of Mitochondrial Haplotypes in Australian *Drosophila melanogaster*. *Mol Biol Evol*, 34(10), 2600-2612. doi:10.1093/molbev/msx184
- Camus, M. F., Wolf, J. B., Morrow, E. H., & Dowling, D. K. (2015). Single Nucleotides in the mtDNA Sequence Modify Mitochondrial Molecular Function and Are Associated with Sex-Specific Effects on Fertility and Aging. *Curr Biol*, 25(20), 2717-2722. doi:10.1016/j.cub.2015.09.012

- Cannon, B., & Nedergaard, J. A. N. (2004). Brown Adipose Tissue: Function and Physiological Significance. *Physiological Reviews*, 84(1), 277-359. doi:10.1152/physrev.00015.2003
- Cardoso, A. R., Kakimoto, P. A. H. B., & Kowaltowski, A. J. (2013). Diet-sensitive sources of reactive oxygen species in liver mitochondria: role of very long chain acyl-CoA dehydrogenases. *PLoS One*, 8(10), e77088-e77088. doi:10.1371/journal.pone.0077088
- Casteels, K., Ong, K., Phillips, D., Bendall, H., & Pembrey, M. (1999). Mitochondrial 16189 variant, thinness at birth, and type-2 diabetes. ALSPAC study team. Avon Longitudinal Study of Pregnancy and Childhood. *Lancet*, 353(9163), 1499-1500.
- Cavalli-Sforza, L. L., & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nature Genetics*, 33, 266. doi:10.1038/ng1113
- Chapman, T. P., Hadley, G., Fratter, C., Cullen, S. N., Bax, B. E., Bain, M. D., . . . Travis, S. P. (2014). Unexplained gastrointestinal symptoms: think mitochondrial disease. *Dig Liver Dis*, 46(1), 1-8. doi:10.1016/j.dld.2013.04.008
- Charlesworth, B., & Charlesworth, D. (2010). *Elements of evolutionary genetics*. Greenwood Village, Colo.: Roberts and Co. Publishers.
- Charoute, H., Kefi, R., Bounaceur, S., Benrahma, H., Reguig, A., Kandil, M., . . . Barakat, A. (2018). Novel variants of mitochondrial DNA associated with Type 2 diabetes mellitus in Moroccan population. *Mitochondrial DNA Part A*, 29(1), 9-13. doi:10.1080/24701394.2016.1233530
- Cooper, G. (2000). The Mechanism of Oxidative Phosphorylation. In *The Cell: A Molecular Approach. 2nd edition*. Sunderland (MA): Sinauer Associates. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK9885/>.
- Cormier, R. P. J., Champigny, C. M., Simard, C. J., St-Coeur, P.-D., & Pichaud, N. (2019). Dynamic mitochondrial responses to a high-fat diet in *Drosophila melanogaster*. *Sci Rep*, 9(1), 4531. doi:10.1038/s41598-018-36060-5
- Dassanayaka, S., Readnower, R. D., Salabei, J. K., Long, B. W., Aird, A. L., Zheng, Y.-T., . . . Jones, S. P. (2015). High glucose induces mitochondrial dysfunction independently of protein O-GlcNAcylation. *The Biochemical journal*, 467(1), 115-126. doi:10.1042/BJ20141018

- De Fanti, S., Vicario, S., Lang, M., Simone, D., Magli, C., Luiselli, D., . . . Romeo, G. (2017). Intra-individual purifying selection on mitochondrial DNA variants during human oogenesis. *Human reproduction (Oxford, England)*, 32(5), 1100-1107. doi:10.1093/humrep/dex051
- Díaz, M., Aragonés, G., Sánchez-Infantes, D., Bassols, J., Pérez-Cruz, M., de Zegher, F., . . . Ibáñez, L. (2014). Mitochondrial DNA in Placenta: Associations with Fetal Growth and Superoxide Dismutase Activity. *Hormone Research in Paediatrics*, 82(5), 303-309. doi:10.1159/000366079
- Dillon, L., Amorevieta-Gentil, M., Caron, M., Lewis, C., Guay-Giroux, A., Desjardins, B., & Gagnon, A. (2018). The Programme de recherche en démographie historique: past, present and future developments in family reconstitution. *The History of the Family*, 23(1), 20-53. doi:10.1080/1081602X.2016.1222501
- Doyon, A., Labuda, D., Moreau, C., Larmuseau, M., Vézina, H., Noël, Y., . . . Milot, E. (soumis). Population-wide assessment of spatiotemporal variation in DNA lineage marker frequencies and implications for their use in forensic science. *Manuscript Draft*.
- Duhig, K., Chappell, L. C., & Shennan, A. H. (2016). Oxidative stress in pregnancy and reproduction. *Obstetric medicine*, 9(3), 113-116. doi:10.1177/1753495X16648495
- Dunger, D. B., Petry, C. J., & Ong, K. K. (2007). Genetics of Size at Birth. *Diabetes Care*, 30(Supplement 2), S150. doi:10.2337/dc07-s208
- Elson, J. L., Turnbull, D. M., & Howell, N. (2004). Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet*, 74(2), 229-238. doi:10.1086/381505
- Ewing, B., & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome research*, 8(3), 186-194.
- Fazzini, F., Schöpf, B., Blatzer, M., Coassin, S., Hicks, A. A., Kronenberg, F., & Fendt, L. (2018). Plasmid-normalized quantification of relative mitochondrial DNA copy number. *Sci Rep*, 8(1), 15347. doi:10.1038/s41598-018-33684-5
- Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133(3), 693-709.
- Gagnon, A., & Mazan, R. (2009). Does exposure to infectious diseases in infancy affect old-age mortality? Evidence from a pre-industrial population. *Social Science & Medicine*, 68(9), 1609-1616. doi:https://doi.org/10.1016/j.socscimed.2009.02.008

- Galtier, N., Nabholz, B., GléMin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol*, 18(22), 4541-4550. doi:10.1111/j.1365-294X.2009.04380.x
- García-Ruiz, I., Solís-Muñoz, P., Fernández-Moreira, D., Grau, M., Colina, F., Muñoz-Yagüe, T., & Solís-Herruzo, J. A. (2014). High-fat diet decreases activity of the oxidative phosphorylation complexes and causes nonalcoholic steatohepatitis in mice. *Disease Models & Mechanisms*, 7(11), 1287. doi:10.1242/dmm.016766
- Garvin, M. R., Bielawski, J. P., Sazanov, L. A., & Gharrett, A. J. (2015). Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *Journal of Zoological Systematics and Evolutionary Research*, 53(1), 1-17. doi:10.1111/jzs.12079
- Gemmell, N. J., Metcalf, V. J., & Allendorf, F. W. (2004). Mother's curse: the effect of mtDNA on individual fitness and population viability. *Trends Ecol Evol*, 19(5), 238-244. doi: 10.1016/j.tree.2004.02.002
- Gillespie, D. O. S., Russell, A. F., & Lummaa, V. (2013). The effect of maternal age and reproductive history on offspring survival and lifetime reproduction in preindustrial humans. *Evolution*, 67(7), 1964-1974. doi:10.1111/evo.12078
- Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*, 11(5), 725-736. doi:10.1093/oxfordjournals.molbev.a040153
- Gómez-Durán, A., Díez-Sánchez, C., Pacheu-Grau, D., López-Gallardo, E., Montoya, J., López-Pérez, M. J., & Ruiz-Pesini, E. (2010). Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups. *Human Molecular Genetics*, 19(17), 3343-3353. doi:10.1093/hmg/ddq246
- Granger, D. N., & Kvietys, P. R. (2015). Reperfusion injury and reactive oxygen species: The evolution of a concept. *Redox biology*, 6, 524-551. doi:10.1016/j.redox.2015.08.020
- Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics* (4th ed.). Sunderland, Mass.: Sinauer Associates.
- Howell, N. (1998). Leber hereditary optic neuropathy: respiratory chain dysfunction and degeneration of the optic nerve. *Vision Research*, 38(10), 1495-1504. doi:https://doi.org/10.1016/S0042-6989(97)00444-6

- Hung, C. M., & Zink, R. M. (2014). Distinguishing the effects of selection from demographic history in the genetic variation of two sister passerines based on mitochondrial-nuclear comparison. *Heredity (Edinb)*, 113(1), 42-51. doi:10.1038/hdy.2014.9
- Jastroch, M., Divakaruni, A. S., Mookerjee, S., Treberg, J. R., & Brand, M. D. (2010). Mitochondrial proton and electron leaks. *Essays in biochemistry*, 47, 53-67. doi:10.1042/bse0470053
- Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS genetics*, 4(12), e1000304-e1000304. doi:10.1371/journal.pgen.1000304
- Kurland, C. G., & Andersson, S. G. (2000). Origin and evolution of the mitochondrial proteome. *Microbiology and molecular biology reviews: MMBR*, 64(4), 786-820. doi:10.1128/mmbr.64.4.786-820.2000
- Larsen, S., Diez-Sanchez, C., Rabol, R., Ara, I., Dela, F., & Helge, J. W. (2014). Increased intrinsic mitochondrial function in humans with mitochondrial haplogroup H. *Biochim Biophys Acta*, 1837(2), 226-231. doi:10.1016/j.bbabbio.2013.10.009
- Levine, D. (1976). The reliability of parochial registration and the representativeness of family reconstitution. *Popul Stud (Camb)*, 30(1), 107-122. doi:10.1080/00324728.1976.10412723
- Lodish, H. F. (1997). *Biologie moléculaire de la cellule* (1re éd. ed.). Bruxelles: De Boeck.
- Lowe, I., Bize, P., Lehto Hürlimann, M., & Heckel, G. (2018). Effects of the Mitochondrial and Nuclear Genomes on Nonshivering Thermogenesis in a Wild Derived Rodent. *Integrative and Comparative Biology*, 58(3), 532-543. doi:10.1093/icb/icy072
- Lunze, K., Bloom, D. E., Jamison, D. T., & Hamer, D. H. (2013). The global burden of neonatal hypothermia: systematic review of a major challenge for newborn survival. *BMC medicine*, 11, 24-24. doi:10.1186/1741-7015-11-24
- Lutz-Bonengel, S., & Parson, W. (2019). No further evidence for paternal leakage of mitochondrial DNA in humans yet. *Proceedings of the National Academy of Sciences*, 116(6), 1821. doi:10.1073/pnas.1820533116
- Malik, S., Diot, A., Morten, K., Dombi, E., Vatish, M., Boyd, C. A. R., & Poulton, J. (2017). Acute nutritional stress during pregnancy affects placental efficiency, fetal growth and adult glucose homeostasis. *Oncotarget*, 8(65), 109478-109486. doi:10.18632/oncotarget.22695

- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351(6328), 652-654. doi:10.1038/351652a0
- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Wallace, D. C. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A*, 100(1), 171-176. doi:10.1073/pnas.0136972100
- Moilanen, J. S., Finnila, S., & Majamaa, K. (2003). Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup J. *Mol Biol Evol*, 20(12), 2132-2142. doi:10.1093/molbev/msg230
- Moorad, J. A. (2014). Individual fitness and phenotypic selection in age-structured populations with constant growth rates. *Ecology*, 95(4), 1087-1095.
- Moore, V. M., Davies, M. J., Willson, K. J., Worsley, A., & Robinson, J. S. (2004). Dietary composition of pregnant women is related to size of the baby at birth. *J Nutr*, 134(7), 1820-1826. doi:10.1093/jn/134.7.1820
- Moreau, C., Bherer, C., Vezina, H., Jomphe, M., Labuda, D., & Excoffier, L. (2011). Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science*, 334(6059), 1148-1150. doi:10.1126/science.1212880
- Moreau, C., Vézina, H., Jomphe, M., Lavoie, È.-M., Roy-Gagnon, M.-H., & Labuda, D. (2011). When Genetics and Genealogies Tell Different Stories—Maternal Lineages in Gaspesia. *Ann Hum Genet*, 75(2), 247-254. doi:10.1111/j.1469-1809.2010.00617.x
- Moreau, C., Vezina, H., Yotova, V., Hamon, R., de Knijff, P., Sinnett, D., & Labuda, D. (2009). Genetic heterogeneity in regional populations of Quebec-parental lineages in the Gaspé Peninsula. *Am J Phys Anthropol*, 139(4), 512-522. doi:10.1002/ajpa.21012
- Mullany, L. C. (2010). Neonatal hypothermia in low-resource settings. *Seminars in perinatology*, 34(6), 426-433. doi:10.1053/j.semperi.2010.09.007
- Murphy, M. P. (2009). How mitochondria produce reactive oxygen species. *The Biochemical journal*, 417(1), 1-13. doi:10.1042/BJ20081386
- Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*, 3(5), 418-426. doi:10.1093/oxfordjournals.molbev.a040410

- Otten, A. B., & Smeets, H. J. (2015). Evolutionary defined role of the mitochondrial DNA in fertility, disease and ageing. *Hum Reprod Update*, 21(5), 671-689. doi:10.1093/humupd/dmv024
- Ounsted, M., Scott, A., & Ounsted, C. (1986). Transmission through the female line of a mechanism constraining human fetal growth. *Annals of Human Biology*, 13(2), 143-151. doi:10.1080/03014468600008281
- Pichaud, N., Messmer, M., Correa, C. C., & Ballard, J. W. O. (2013). Diet influences the intake target and mitochondrial functions of *Drosophila melanogaster* males. *Mitochondrion*, 13(6), 817-822. doi:https://doi.org/10.1016/j.mito.2013.05.008
- Ramírez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F., & Navarro, A. (2008). Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*, 179(1), 555-567. doi:10.1534/genetics.107.083006
- Ridge, P. G., & Kauwe, J. S. K. (2018). Mitochondria and Alzheimer's Disease: the Role of Mitochondrial Genetic Variation. *Current genetic medicine reports*, 6(1), 1-10. doi:10.1007/s40142-018-0132-2
- Rinninella, E., Pizzoferrato, M., Cintoni, M., Servidei, S., & Mele, M. C. (2018). Nutritional support in mitochondrial diseases: the state of the art. *Eur Rev Med Pharmacol Sci*, 22(13), 4288-4298. doi:10.26355/eurrev_201807_15425
- Rousset, S., Alves-Guerra, M.-C., Mozo, J., Miroux, B., Cassard-Doulcier, A.-M., Bouillaud, F., & Ricquier, D. (2004). The Biology of Mitochondrial Uncoupling Proteins. *Diabetes*, 53(suppl 1), S130. doi:10.2337/diabetes.53.2007.S130
- Saneto, R. P., & Sedensky, M. M. (2013). Mitochondrial disease in childhood: mtDNA encoded. *Neurotherapeutics: the journal of the American Society for Experimental NeuroTherapeutics*, 10(2), 199-211. doi:10.1007/s13311-012-0167-0
- Santos, C., Montiel, R., Arruda, A., Alvarez, L., Aluja, M. P., & Lima, M. (2008). Mutation patterns of mtDNA: empirical inferences for the coding region. *BMC Evol Biol*, 8, 167-167. doi:10.1186/1471-2148-8-167
- Sanz, A., Caro, P., Ayala, V., Portero-Otin, M., Pamplona, R., & Barja, G. (2006). Methionine restriction decreases mitochondrial oxygen radical generation and leak as well as oxidative damage to mitochondrial DNA and proteins. *The FASEB Journal*, 20(8), 1064-1073. doi:10.1096/fj.05-5568com

- Selman, C., Villasevil, E. M., Anderson, G. J., Metcalfe, N. B., Salin, K., & Chinopoulos, C. (2018). The RCR and ATP/O Indices Can Give Contradictory Messages about Mitochondrial Efficiency. *Integrative and Comparative Biology*, 58(3), 486-494. doi:10.1093/icb/icy085
- Singh, K. K., & Kulawiec, M. (2009). Mitochondrial DNA polymorphism and risk of cancer. *Methods in molecular biology (Clifton, N.J.)*, 471, 291-303. doi:10.1007/978-1-59745-416-2_15
- Skjærvø, G. R., Fossøy, F., & Røskoft, E. (2015). Solar activity at birth predicted infant survival and women's fertility in historical Norway. *Proc Biol Sci*, 282(1801), 20142032-20142032. doi:10.1098/rspb.2014.2032
- Song, S., Pursell, Z. F., Copeland, W. C., Longley, M. J., Kunkel, T. A., & Mathews, C. K. (2005). DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. *Proc Natl Acad Sci U S A*, 102(14), 4990-4995. doi:10.1073/pnas.0500253102
- Sorg, M., & Craig, B. (1983). Patterns of Infant Mortality in the Upper St. John Valley French Population: 1791-1838. *Human biology*, 55, 100-113.
- Stajich, J. E., & Hahn, M. W. (2005). Disentangling the effects of demography and selection in human history. *Mol Biol Evol*, 22(1), 63-73. doi:10.1093/molbev/msh252
- Subramanian, S. (2016). The effects of sample size on population genomic analyses – implications for the tests of neutrality. *BMC Genomics*, 17(1), 123. doi:10.1186/s12864-016-2441-8
- Sun, C., Kong, Q.-P., & Zhang, Y.-P. (2007). The role of climate in human mitochondrial DNA evolution: A reappraisal. *Genomics*, 89(3), 338-342. doi:https://doi.org/10.1016/j.ygeno.2006.11.005
- Sun, X., Wheeler, C. T., Yolitz, J., Laslo, M., Alberico, T., Sun, Y., Zou, S. (2014). A mitochondrial ATP synthase subunit interacts with TOR signaling to modulate protein homeostasis and lifespan in Drosophila. *Cell reports*, 8(6), 1781-1792. doi:10.1016/j.celrep.2014.08.022
- Swerdlow, R. H. (2012). Does mitochondrial DNA play a role in Parkinson's disease? A review of cybrid and other supportive evidence. *Antioxidants & redox signaling*, 16(9), 950-964. doi:10.1089/ars.2011.3948

- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585-595.
- Tarín, J. J., Brines, J., & Cano, A. (1998). Long-term effects of delayed parenthood. *Human Reproduction*, 13(9), 2371-2376. doi:10.1093/humrep/13.9.2371
- Templeton, A. R. (2006). *Population genetics and microevolutionary theory*. In (pp. 1 ressource en ligne). Retrieved from <http://onlinelibrary.wiley.com/book/10.1002/0470047356>
- Towarnicki, S. G., & Ballard, J. W. O. (2017). Drosophila mitotypes determine developmental time in a diet and temperature dependent manner. *Journal of Insect Physiology*, 100, 133-139. doi:<https://doi.org/10.1016/j.jinsphys.2017.06.002>
- Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu Rev Genet*, 47, 97-120. doi:10.1146/annurev-genet-111212-133526
- Vives-Bauza, C., Gonzalo, R., Manfredi, G., Garcia-Arumi, E., & Andreu, A. L. (2006). Enhanced ROS production and antioxidant defenses in cybrids harbouring mutations in mtDNA. *Neuroscience Letters*, 391(3), 136-141. doi:<https://doi.org/10.1016/j.neulet.2005.08.049>
- Wallace, D. C. (2015). Mitochondrial DNA variation in human radiation and disease. *Cell*, 163(1), 33-38. doi:10.1016/j.cell.2015.08.067
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., . . . Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, 44(W1), W58-W63. doi:10.1093/nar/gkw233
- Yang, Z., & Nielsen, R. (2000). Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol Biol Evol*, 17(1), 32-43. doi:10.1093/oxfordjournals.molbev.a026236
- Zhang, J., Nuebel, E., Wisidagama, D. R. R., Setoguchi, K., Hong, J. S., Van Horn, C. M., . . . Teitell, M. A. (2012). Measuring energy metabolism in cultured cells, including human pluripotent stem cells and differentiated cells. *Nature protocols*, 7(6), 1068-1085. doi:10.1038/nprot.2012.048